Center for the Mathematics of Uncertainty

An Introduction to the Mathematics of Uncertainty

including
Set Theory, Logic,
Probability, Fuzzy Sets,
Rough Sets, *and* Evidence Theory

Mark J. Wierman

August 20, 2010

Honors Program
Creighton
UNIVERSITY
College of Arts
and Sciences

Sponsoring Editor: *Creighton University*

To my wife
Mary Dobransky

Contents

List of Tables

List of Figures

List of Algorithms

Nomenclature

$\neg$	logical negation textsfnot
$\rightarrow$	logical implication if–then
$\vee$	logical disjunction or
$\wedge$	logical conjunction and
(a, b)	Open interval from a to b
$[a, b]$	Closed interval from a to b
χ_A	Characteristic function of A
$\circ_{\mathsf{t}}$	sup–t composition of fuzzy relations
$\emptyset$	The empty set
$\exists$	universal quantifier there exists
$\forall$	universal quantifier for all
$\frac{df}{dx}$	The derivative of f
A^{α}	Alpha–cut of A
$\mathsf{c}(a)$	Complement function
$\mathsf{h}(a, b)$	Averaging function, binary
$\mathsf{h}(a_1, a_2, \ldots, a_n)$	Averaging function, general
$\in$	Element of
$[0, 1]$	The unit interval
$\langle x, y \rangle$	Ordered pair of x and y
$\mathbb{N}_n$	The natural numbers up to n
$\mathcal{P}(X)$	The power set of X
q	possibility distribution
$\mu_{\mathsf{A}}(x)$	The membership function of a fuzzy set
$\omega_{\mathsf{t}}(a, b)$	Residuum function
$\subseteq$	subset or equal
i_a	Lukasiewicz
i_{gg}	Goguen
i_{gr}	Gaines–Rescher
i_g	Gödel
i_{kd}	Kleene-Dienes
i_l	Larsen
i_m	Early Zadeh
i_r	Reichenbach strict
i_{ss}	Standard strict
i_y	Yager
$\times$	Cartesiona product of sets
$C(\mathsf{A})$	Core of A
$dist(a, b)$	Distance between points
$dist_D$	Displacement distance
$dist_E$	Euclidean distance
$dist_H$	Hamming distance

$dist_M$	Minkowski distance
$E(X)$	Expected value of X
$f \circ g$	The composition of functions
$f(A)$	The extension of function f to a set A
$f(x)$	A function f of x
$f^{(-1)}$	Pseudo–inverse function of f
f^{-1}	Inverse function of f
f^{-1}	The inverse of a function f
$f'(x)$	The derivative of f
$h(\mathsf{A})$	Height of A
$I(\mathsf{A})$	Image set of A
L	Lattice
$OWA_{\mathbf{w}}$	Ordered weighted average
P	Probability measure
p	Probability distribution
$p(\mathsf{A})$	Plinth of A
Q	Possibility measre
$S(\mathsf{A})$	Support of A
X, Y, Z	Universal sets
DS	Decision system.
FLC	Fuzzy Logic Controller
$\mathsf{A}^{\alpha+}$	Strong alpha–cut of A
$\mathbb{B}$	The Boole set $\{0, 1\}$.
A^{c}	Fuzzy complement of A, usually $1 - \mathsf{A}(x)$
$\mathsf{A} \setminus \mathsf{B}$	Fuzzy set difference
$Dist_{C_\infty}(\mathsf{A}, \mathsf{B})$	Maximal vertical distance between fuzzy numbers
$Dist_H(\mathsf{A}, \mathsf{B})$	Hamming distance between fuzzy numbers
$Dist_D(\mathsf{A}, \mathsf{B})$	Hausdorf distance between fuzzy numbers
m	Basic probability assignment
$\mathsf{Tb}\langle m, s, \gamma \rangle$	Guassian bell fuzzy number
$\mathsf{Ti}\langle m \rangle$	Impulse fuzzy number
$\mathsf{Tq}\langle a, l, m, r, b \rangle$	Piecewise quadratic fuzzy number fuzzy number
Sld	S–shaped linear decreasing fuzzy number
Sli	S–shaped linear increasing fuzzy number
Sqd	S–shaped quadratic decreasing fuzzy number
Sqi	S–shaped quadratic increasing fuzzy number
$\mathsf{S}\sigma$	S–shaped sigmoid fuzzy number
S	S–shaped fuzzy number
$\mathsf{Tp}\langle a, l, r, b \rangle$	Trapezoidal fuzzy number (fuzzy interval)
$\mathsf{Tr}\langle a, m, b \rangle$	Triangular fuzzy number
Tlr	Left–right, or $L - R$, fuzzy number
A, B, C	The label or tags of a fuzzy sets
i	Fuzzy implication operator
$\wedge$	Fuzzy intersection, usually $\min$

$\mu_X^R(x)$	RST: rough set membership function.
$\frac{1.0}{a} + \frac{0.7}{b}$	Fraction notation for fuzzy set membership function
$\mathsf{A}(x)$	The membership functions of a fuzzy set
$\mathbb{Z}$	The integers
$\mathbb{N}$	The natural numbers
$\mathbb{R}$	The real numbers
$\oplus$	Fuzzy or interval addition
$\oslash$	Fuzzy or interval division
A^{-1}	Fuzzy or interval inverse
$\otimes$	Fuzzy or interval multiplication
$\ominus$	Fuzzy or interval subtraction
p	probabiltiy distribution
$\mathcal{F}(X)$	Fuzzy power set
$proj_1(C)$	Projection of C into the first dimension X of Z.
$proj_2(C)$	Projection of C into the second dimension Y of Z.
$a\ R\ b$	Relation
$R(x)$	RST: The equivalence class of x under the realtion R.
A^{c}	Set complement of A
$B \setminus A$	Set B minus set A
$A \cap B$	Set intersection
$A \triangle B$	Symmetric difference of A and B
$A \cup B$	Set union
$\{a, b, c\}$	A set of objects
$\subseteq$	Subset
$\mathsf{t}(a, b)$	t-norm intersection function
$\mathsf{s}(a, b)$	t-conorm union function
$\vee$	Fuzzy union, usually $\max$
$\mathbb{U}$	The unit interval $[0, 1]$.

Greek alphabet

Letters		*Names*	*Letters*		*Names*	*Letters*		*Names*
A	α	Alpha	I	ι	Iota	P	ρ	Rho
B	β	Beta	K	κ	Kappa	Σ	σ	Sigma
Γ	γ	Gamma	Λ	λ	Lambda	T	τ	Tau
Δ	δ	Delta	M	μ	Mu	Υ	υ	Upsilon
E	ε	Epsilon	N	ν	Nu	Φ	ϕ	Phi
Z	ζ	Zeta	Ξ	ξ	Xi	X	χ	Chi
M	η	Eta	O	o	Omicron	Ψ	ψ	Psi
Θ	θ	Theta	Π	π	Pi	Ω	ω	Omega

Fuzzy Sets and Fuzzy Logic

English	**Logical**		**Set**		**Fuzzy**
not A	negation	$\neg A$ or $\bar{A}$	A^c	complement	$1 - a$
A and B	conjunction	$A \wedge B$	$A \cap B$	intersection	$\min(a, b)$
A or B	disjunction	$A \vee B$	$A \cup B$	union	$\max(a, b)$
if A then B	implication	$A \rightarrow B$	$A \subseteq B$	subsethood	$\begin{cases} 1 & b \geq a \\ b & a > b \end{cases}$

Preface

In 1988, I took a class in Fuzzy Set theory from George Klir at Binghamton University-SUNY. I had a deep background in Set Theory and thought Fuzzy Set Theory would be a good niche subject. In 1992 as I was starting my thesis I owned just about every major book on Fuzzy Set theory and they did not fill up even one bookshelf. It seemed that it would be easy to keep up on all the literature and generate a lot of papers, since it was obvious to me there was a lot of work to be done.

Then came the big success of the “killer app" — fuzzy control. Fuzzy control is a major technology of the modern world, it pilots the space shuttle and its why Japanese cars were smoother to drive.

Fuzzy Set Theory exploded, there are now sections of books on Fuzzy Set Theory at bookstores like Borders and Barnes and Nobles. I was even hired at Creighton University for my expertise in this subject, not my extensive backgrounds in modeling and simulation, which I thought would be my strong point.

Why Fuzzy Set Theory

It is interesting to note that the area of research pursued by Lotfi Zadeh before his creation of Fuzzy Set Theory was adaptive filters, the work which culminated in the Kalman filter. The Kalman filter is an adaptive statistical method. To give an simplistic illustration of an adaptive method, suppose that we know that the average of ten grades is 85.3 and that a student receives a 88 on his eleventh test. It is easy to show that we can find the new average of all eleven tests using the formula $\frac{85.3*10+88}{11}$ which shows that we don’t need the first ten values to compute the current average. Similarly the Kalman filter constantly updates values used in filtering input data or controlling a mechanical device.

Filtering input data, hearing a single voice say your name in a room full of conversations, or controlling a mechanical device, driving a car, are skills that the human brain is extremely good at. Computers, even with techniques such as Kalman filters have difficulty with similar skills whenever the systems are complicated, such as continuous speech recognition or controlling a helicopter’s flight. It would seem that this dichotomy led Zadeh to develop a method that mirrors the way that humans think. How do we think? Loosely yet powerfully.

Humans make statements to one another such as, “*Can you get me the blue shirt?*” that are extremely vague. Its not a question, its a request or possibly an order. The shirt requested may be any style of shirt, button-down, polo, t-shirt, short or long sleeved. The shade of blue – royal, navy, sea, robin’s-egg, etc., is the vaguest information of all. However, it is exactly these simple short vague statements that demonstrate the enormous amount of information humans can convey to one another with simple statements in their languages.

Here is another example of the power of language; it is much easier to absorb the basic ideas about a topic of interest by attending a lecture rather than reading a book. A good lecture is adaptive in nature, the speaker can elucidate any sticking points and gloss over portions that are in the audiences knowledge domain.

I recommend that initiates to fuzzy logic consult the FAQ for fuzzy logic is at:

http://www.cs.cmu.edu/Web/Groups/AI/html/faqs/ai/fuzzy/part1/faq.html

An example of the information contained there is the following.

> **[8] Isn't "fuzzy logic" an inherent contradiction? Why would anyone want to fuzzify logic?**
>
> Date: 15-APR-93
>
> Fuzzy sets and logic must be viewed as a formal mathematical theory for the representation of uncertainty. Uncertainty is crucial for the management of real systems: if you had to park your car PRECISELY in one place, it would not be possible. Instead, you work within, say, 10 cm tolerances. The presence of uncertainty is the price you pay for handling a complex system.
>
> Nevertheless, fuzzy logic is a mathematical formalism, and a membership grade is a precise number. What's crucial to realize is that fuzzy logic is a logic OF fuzziness, not a logic which is ITSELF fuzzy. But that's OK: just as the laws of probability are not random, so the laws of fuzziness are not vague.

No one has a problem understanding that probability deals with random phenomena. This does not mean that probability theory is itself random. It is the theory of random events. If it was called "random theory" would people think that its formulas depended on the roll of a dice?

Yet the word fuzzy in fuzzy set theory provokes criticism! Fuzzy sets are a precise mathematical tool. The represent information that may be subjective, inadequately described or detailed, aggregated loosely, or ill-understood. We then use the techniques of fuzzy sets/fuzzy logic to make whatever conclusions we can with this data. These conclusions follow a determined mathematical course to a conclusion. However, the conclusions inherit the uncertainty of the premises, and this uncertainty may aggregate, however, as in real life, it may be the best answer we can get.

The need for something like fuzzy set theory is indicated by the the large number of systems that crisp (that is non-fuzzy) mathematical techniques have failed to provide effective methods to deal with. We do not understand the economy, the weather, human emotion, vision, and many other phenomena.

This book provides a fairly narrow trail through the cybernetic landscape. It does not provide all the details of fuzzy sets and fuzzy logic. What it does do is give a fast track to some of the most important fuzzy applications.

Introduction

Consider a coin in your pocket. What is it worth? If the coin is a recently minted quarter then it is probably worth face-value, 25¢. If this coin is older then it may have greater monetary value to a collector. The value of the coin would then depend upon its rarity and its condition.

Let us take the coin out of our pocket and consider its condition. Since it is in you pocket we cannot classify the coin as uncirculated. If it is old it is likely to be worn to some extent. Usually, one would consult an expert (a numismatist) who would try to classify its condition. This classification is somewhat subjective, and can vary between experts. That is because the act of classification, of assigning a tag to an object, is fuzzy.

Fuzzy set theory is the study of just this type of uncertainty, fuzzy sets use numbers to quantify the degree to which a property can be associated with an object. There is nothing fuzzy about a fuzzy set just as there is nothing random about a probability distribution. Suppose we are now looking at the uncertainty associated with the act of flipping the coin. Suppose that it is stated that the probability distribution for the flipping the coin event is $p = \left\{\frac{1}{2}, \frac{1}{2}\right\}$ for $X = \{H, T\}$. This statement says that the coin, when flipped, turns up as either $H = heads$ or $T = tails$. The probability statement also contends that long term results of repeatedly flipping the coin produces a ratio of heads to total flips that is about $\frac{1}{2}$ and of tails to total flips that is also around $\frac{1}{2}$. Probability theory is the study of random events. The numbers come from random trials such as the flipping of a coin.

This is very nice information but it is totally irrelevant to the question, “What is this coin worth?” We could poll many expert numismatists with random coins to determine the probability that a coin is worth 2 ¢ but that would still not answer the question, “What is this coin worth?”

Fuzzy set theory is the study of fuzzy events. A fuzzy event is one that is difficult to classify, like the flavor of a apple. Fuzzy set theory is the study of vagueness. The statement “We will meet this evening” is vague. This evening is not a specific time. There is not any firm boundary between evening and night and this is the type of uncertainty that a fuzzy set theory is constructed to represent and process.

If one asks a variety of experts, “On a scale of one to ten, to what degree do you consider this coin as being in Good condition?” one might find that the average of the experts opinion, divided by ten is 0.73. If one then asks the same group of experts, “On a scale of one to ten, to what degree do you consider this coin as being in Fair condition?” one might find that the average of the experts opinion, divided by ten is 0.42. This kind of experiment can easily be carried out in real life. Note that there is nothing random about the individual answers, repetition will tend to produce identical results. Note also that the results do not need to sum up to one. Human evaluation is often inexact, i.e., fuzzy. An expert could change his mind right in the middle of a series of questions.

The bibliography references at the end of each chapter are not intended to be comprehensive. The bibliography includes mostly references to the originators of important theoretical ideas as well as the best books, in the authors opinion, to delve deeper into the applications presented.

Part I.

Uncertainty

1. Uncertainty

Uncertainty is a universal dilemma. Uncertainty intrudes into plans for the future, interpretations of the past, and decisions in the present.

There are many kinds of uncertainty.

Suppose it is our task to study a Dead Sea Scroll. Let us consider some of the uncertainties involved in the analysis of the scroll. These include:

1. the precise date of composition,
2. the author's identity,
3. where was it composed,
4. the origin and composition of the ink and parchment,
5. reading the faded text, and
6. translating the original meaning.

1.1. Example of Different Types of Uncertainty in One Context

As another example, let us suppose that the label on our prescription of Piperol has fallen off. We try to obtain some expert advice from various sources and arrive at the following list Krause and Clark (1993).

Expert	Prescribed Treatment
A	100mg, once every 12 hours
B	80mg-100mg, twice a day
C	About 120mg, 2–3 times a day
D	Likely to be 200mg twice a day
E	30mg? or 80mg? twice a day (the number is hard to read)
F	200mg 4 times a day or 100mg once a day
G	150mg
H	At least 100mg, twice a day
I	The usual dose for this drug is 100mg, twice a day
J	1g, twice a day
K	Google it
L	Never heard of that drug
M	1313 Mockingbird Lane

1.1.1. What Types of Uncertainty?

Let us examine the types of uncertainty contained in each statement.

A 100mg, once every 12 hours

There is very little uncertainty in this statement. This is a *precise* prescription. The quantity of Piperol to take is 100mg. We are to take it two times each day. There is some uncertainty as to when the patient should take the first dose, maybe at 9AM? But after that, every twelve hours the patient should take another dose. But the uncertainty of the starting time is not part of the prescription, it is part of the execution. It is difficult to implementation these instructions incorrectly.

B 80mg-100mg, twice a day

This is an imprecise statement because 80mg–100mg is an interval. Intervals are inherently *vague* because the precise amount of the drug we need to take is not specific. How are we to decide which value in the interval is the best value. This type of instruction is common when the drug in question is available in different strengths (dosages). It is up to the patient to ingest enough, but not too much, and to do that twice each day.

C About 120mg, 2–3 times a day

We know how much of the drug to take but not the frequency. Note that 2–3 is not an interval, we can take the dosage either twice or thrice a day, but not any value in between. This is a *fuzzy* number. This type of prescription is common for painkillers. Since a patients response to pain is very idiosyncratic, some patients will need more, and other patients need less, of the drug to manage the pain.

D Likely to be 200mg twice a day

This is a statement of *confidence*. Most of the time Piperol is prescribed, it is for Beamer's Syndrome. The Physicians Handbook recommends 500mg twice for a patient diagnosed with Beamer's Syndrome. However, sometimes Piperol is prescribed for other ailments. In these cases the dosage varies greatly. Note that this anecdote includes a dosage twice that of the previous instructions.

E 30mg? or 80mg? twice a day (the number is hard to read)

In this case it is hard to determine if the first digit is a 3 or an 8. The printing is smudged. Here the expert tried for precision, but the medium of transmission has failed to correctly convey the information to the patient. The result is *ambiguity*.

F 200mg 4 times a day or 100mg once a day

Here the alternatives are *inconsistent*. One would think that it would be the other way around, 100mg twice a day or 200mg once a day, and that the expert misspoke. Many drugs are powerful and taking too much or too little can be very dangerous. The uncertainty here is very different from that of the previous example. The patient can determine exactly what the instructions say to do, but those instructions appear to be contradictory.

G 150mg

Here we have *incomplete* information. A good prescription tells us every thing we need to know. How much of the drug to take. How to take the drug. When to take the drug. Do we take the drug with food? What things should we avoid when taking the drug. In this case we know how much of the drug to take, but not the schedule. Maybe we only need to take Piperol once, though this is unusual for most drugs.

H At least 100mg, twice a day

This is very *imprecise*. Is this a minimal dose? And if we take a minimal dose will it be effective? What is the optimal dose? What is the maximum dose? The pills are 80mg each? Do the pills break into pieces easily? I hate doing math!

I The usual dose for this drug is 100mg, twice a day

The implied question is: "Does the patient have the usual disease?" This prescription is a default rule, and this rule may or may not be relevant to the patients situation. This information is too *general*.

J 1g, twice a day

Is this a typo? It is *anomalous* when compared to other responses on the list. It is 5 to 10 times as much of the drug as recommended by any of the previous experts. Note that this prescription is precise, and if it were the only instructions available we would have little reason to doubt it. It is only when we compare it to the other prescriptions that we see that it is inconsistent.

K Google it

This may be good advice, in general, but an expert is suppose to provide an answer, not a methodology. This expert avoids the question. This is *incongruency*. It is not like any of the previous answers and, in fact, does not tell us what we need to know. This is especially true since the Internet is not a reliable source.

L Never heard of that drug

Here we have an expert who, it turns out, is not really not an expert in the field. It turns out that this expert is *ignorant* of the uses of the wonder drug Piperol.

M 1313 Mockingbird Lane

Here is answer that is completely *irrelevant*. Maybe the expert misheard the question. Maybe they are an expert in TV trivia and not on drug usage. This is information of doubtful utility.

1.2. Uncertainty Typology

In the previous section we examined data that possessed or displayed some of the following types of uncertainty: *vague, fuzzy, confidence, ambiguity, inconsistent, incomplete, imprecise, general, anomalous, incongruent, ignorant,* and *irrelevant*. While a dictionary definition of uncertainty is unlikely to help us understand the aspects of uncertainty, a good thesaurus can provide an interesting list of synonyms.

Thesaurus

Main Entry:	uncertainty
Part of Speech:	noun
Definition:	doubt, changeableness
Synonyms:	ambiguity, ambivalence, anxiety, bewilderment, concern, confusion, conjecture, contingency, dilemma, disquiet, distrust, doubtfulness, dubiety, guesswork, hesitancy, hesitation, incertitude, inconclusiveness, indecision, irresolution, lack of confidence, misgiving, mistrust, mystification, oscillation, perplexity, puzzle, puzzlement, qualm, quandary, query, questionableness, reserve, scruple, skepticism, suspicion, trouble, uneasiness, unpredictability, vagueness, wonder, worry
Antonyms:	certainty, definiteness, security, sureness

Roget's 21st Century Thesaurus, Third Edition

The thesaurus contains synonyms that were examined in the examples above, like *ambiguity*, while other synonyms, such as *mystification*, were not.

There certainly are a lot of different types of uncertainty. Maybe it would help us if we could organize uncertainty into classifications. Certainly *ambivalence* and *oscillation* are similar in nature, as are *worry* and *anxiety*.

Many people have tried to organize and classify the many types of uncertainty. The typologies often reflect the focus of the investigator's research. Let us take a quick look at four viewpoints.

1.2.1. Typology of Morgan and Henrion – Risk management

Risk management is most prominent in economics (investing in stocks and bond is always risky) and policy planning (how many fireman and police are needed to minimize the risk of disaster). The typology of Morgan and Henrion Morgan and Henrion (1990) (See Table 1.1) is especially useful as it contains aspects best dealt with probabilistically (Random error and statistical variation) as well as aspects that are best modeled with fuzzy set theory (Linguistic imprecision). Here is a brief explication of the uncertainties in the Morgan and Henrion typology.

- Random error and statistical variation occurs whenever we measure something, like the speed of a car.
- Systematic error and subjective judgment occur because the measurement devices themselves are not perfect, and are not always used correctly.

1	Random error and statistical variation
2	Systematic error and subjective judgment
3	Linguistic imprecision
4	Variability
5	Randomness and unpredictability
6	Expert Uncertainty
7	Approximation
8	Model uncertainty

Table 1.1.: Morgan and Henrion

- Linguistic imprecision occurs when we say the car is fast, what velocity is fast?
- Variability in a population is common, seemingly identical cars will not perform exactly the same on a speedway.
- Randomness and unpredictability are seen when a car fails to live to go as fast as it was designed to go.
- Expert Uncertainty is seen when a car is rated by consumer websites, which often give disparate valuations to the same model car.
- Approximation occurs when we round a velocity to the nearest mile per hour.
- Model uncertainty occurs when we model a cars expected lifetime based on test results.

1.2.2. Smithson Typology–Behavioral Science

Smithson's typology Smithson (1989, 1990) comes from the behavioral sciences. In it (See Figure (1.1)) ignorance is the root and uncertainty is just one of many problems in planning for future disasters. Personally, I think it is missing a branch for Stupidity.

1.2.3. Klir Typology – General Systems

In the latter half of the 20th century computer technology impacted every area of technology. One of the most important results of computer technology was the recognition of the inherent similarity of many different systems. For example, plumbing houses very similar to wiring up chips on a circuit board . Houses and apartments correspond to chips and resistors. Wires correspond to pipes, and there are rules to ensure minimal cost, adequate capacity, and separation (to prevent interference/contamination).

The development of measures of uncertainty in mathematical systems has been a major component of Klir's research Klir and Wierman (1998). This has led him to classify uncertainty into two major categories, fuzziness, which deals with information that is indistinct, and ambiguity, which deals with multiplicity. In the first case

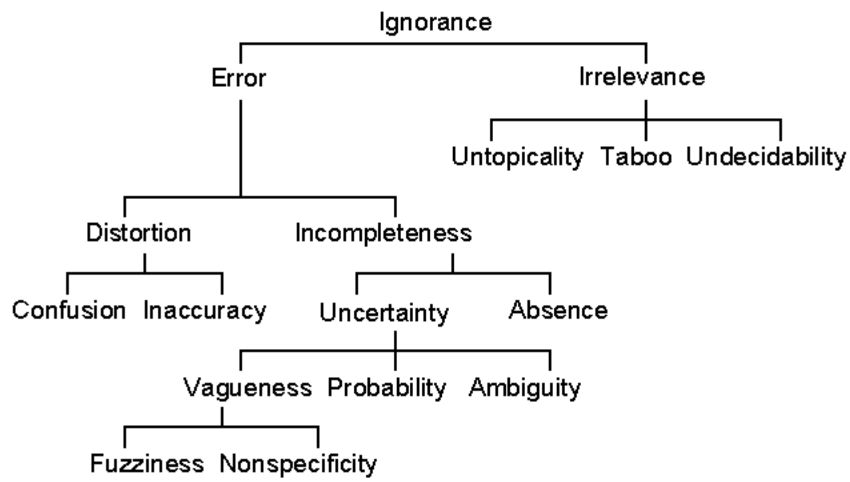


Figure 1.1.: Smithson's Typology.

the telescope sees one object, but does not have the resolution to determine its identity. In the second case the telescope sees multiple objects, and what we see does not allow us to identify precisely the object we were seeking. See Figure (1.2).

1.2.3.1. Potocan, Mulej and Kajzer Typology–Cybernetics

Cybernetics concerns the study of regulatory systems. Common examples of regulatory systems are cruise control in cars, and auto-focus in cameras.

Potocan, Mulej, and Kajzer apply cybernetics and system theory to business systems. It is interesting that their result are presented very differently from Smithson or Klir's, which are presented as tree structures. Instead we have eight aspects of uncertainty, and a system can contain any or all of these characteristics. There is much overlap between Potocan, Mulej, and Kajzer's system with the types of uncertainty previously discussed. However, Pocotan, Mulej and Kajzer introduce some interesting new aspects, like the difference between natural and artificial systems, or the difference between open and closed systems. A spaceship is a closed system that tries to provide a natural environment. Samurai warriors were part of an artificial system which was closed to outsiders.

1.3. Conclusions

There are many types of uncertainty, so it should not be any surprise that there are many different mathematical systems that have been developed to calculate uncertainties. For the purpose of this book, the following are the most important mathematical models of uncertainty.

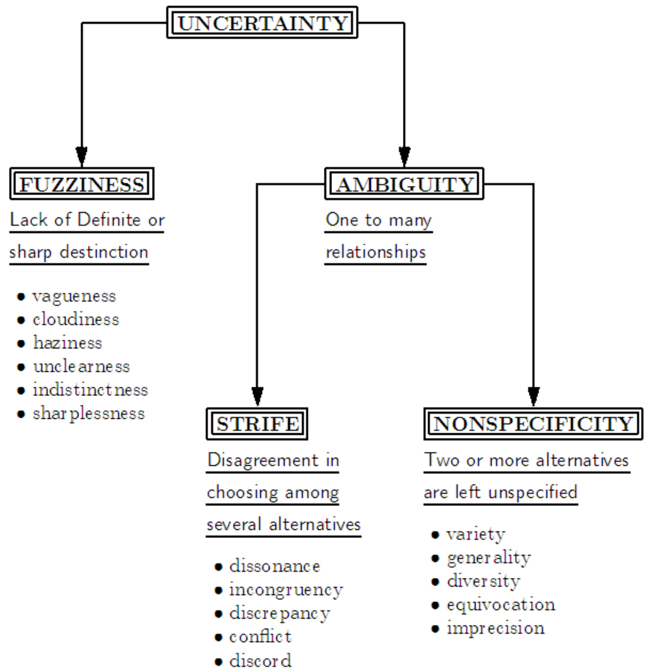


Figure 1.2.: Klir

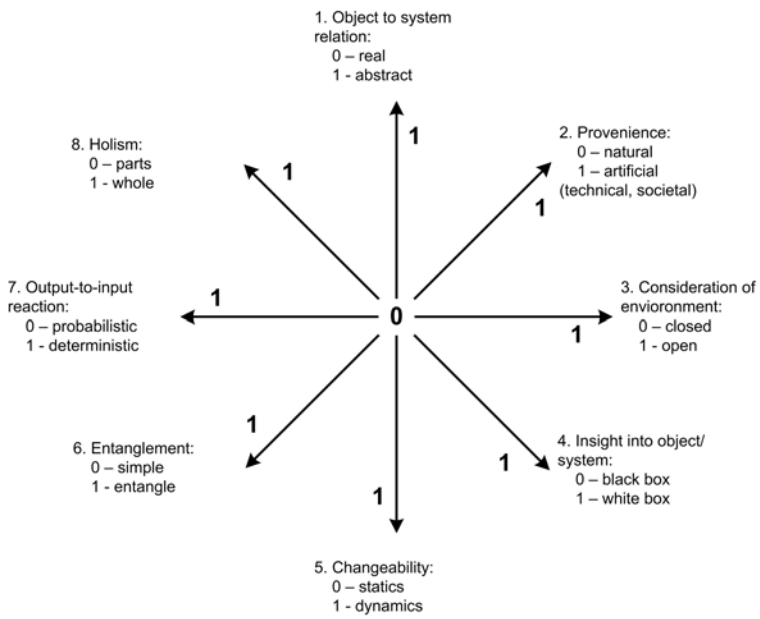


Figure 1.3.: Pocotan.

1. Set theory
2. Probability theory
3. Logic
4. Fuzzy set theory
5. Evidence theory
6. Rough set theory

The second chapter will cover set theory, and the third will provide a brief overview of probability theory and statistics.

Homework

1. What are some of the uncertainties involved in the following quote?

> I think the probability of the Bozo the Clown party winning the next election is between 60 and 70% – Bozo the Clown

1. Use a thesaurus to find five synonyms for uncertainty not examined in Section (1.1). Provide prescriptions that exemplify these synonyms. Explain these prescriptions.
2. What is your primary field of study? Give five examples of the types of uncertainty that commonly occur in your primary field of study. Explain your examples.
3. Use a thesaurus to provide a list of synonyms for uncertainty. Group all of these synonyms into types. Explain your system of classification of uncertainty into types.
4. What is the greatest uncertainty that you faced today? This year? In your life so far?
5. What color pants will you wear tomorrow. Describe any uncertainties associated with this prediction.
6. Without consulting any references, answer this question: “How much exercise should *you* average?”
7. Discuss the uncertainty in a typical conundrum like
   a) How long will the universe last?
   b) Who will win the Superbowl this year?
   c) What are you having for dinner today?
   d) How many angels can dance on the head of a pin?

2. Set Theory

2.1. Sets and uncertainty

The development of set theory and probability theory are intertwined. Early probability theory often focused on games of chance. Probability theory developed methods to answer questions like “what are the odds that I will draw to an inside strait in a hand of poker.” In poker a straight is five cards in sequence. A poker player has a straight draw when he has four out of the five cards of a sequence. If the missing card is in the interior of the sequence then it is called inside straight draw. An example is the poker hand consisting of $3\diamondsuit$, $4\clubsuit$, $6\heartsuit$, $7\spadesuit$, and $J\spadesuit$. The players best strategy is to discard the $J\spadesuit$ and hope for a five of any suit.

A typical poker hand is five cards out of a deck of 52 cards. A specific poker hand is five cards such as $3\diamondsuit$, $4\clubsuit$, $6\heartsuit$, $7\spadesuit$, and $J\spadesuit$. A poker hand is neither a number, nor is it a geometric object. For thousands of years mathematics was concerned itself primarily with numbers and with geometric objects. In addition, logic dealt with True and False.

A specific poker hand is a sub-collection of five objects out of an larger collection of 52 objects. In set theory the universal set, universe of discourse, or just universe, is the collection of objects that is under discussion. All sets can contain any, none, or all of the objects in that universe.

In terms of uncertainty, a universal set is a powerful mechanism. It puts specific limits on what can be discussed, manipulated, and analyzed. If the universal set is a standard deck of 52 cards then the $3\heartsuit$ is admissible but the $Joker$ is not admissible. A universal set sharply limits ambiguity. It allows for the construction of many to one relationships. For example, there are four cards in the deck that can fill the inside straight in the above example, they are the $5\diamondsuit$, $5\clubsuit$, $5\heartsuit$, or $5\spadesuit$.

2.2. Basics

Set theory is the foundation of all branches of modern mathematics. Even numbers and geometric objects are seldom considered as basic or primitive concepts. In formal mathematics they are defined as constructions of set theory.

Set theory is based on the notion of a class or collection of objects. The fundamental concept of set theory is; given an object in the universe of discourse (the general assemblage of things that a discussion is about), a set is *well defined* if one can decide whether or not the given object is contained in the set. This simple notion was one of the most powerful ideas ever conceived in the field of mathematics.

A set is a collection of objects called elements. Typographically the brackets “$\{$” and “$\}$” are used to denote the beginning and ending of the list of elements that are in the set.

A set is defined using one of three methodologies. In the first method the elements of the set are explicitly listed, as in

$$A = \{1, 3, 5, 7, 9\} \tag{2.1}$$

Here we have a set, tagged with the name or label A, and containing as elements the objects one, three, five, seven, and nine. Symbolically the statement "5 is an element of set A" is written $5 \in A$. We can also say that 6 is not an element of A, or $6 \notin A$. Given an object in the universe of discourse we can now compare it to the elements in the list defining the set A. If there is a match then the object is in the set. If there is no match then the element is not in the set. Conventionally capital letters represent sets and small letters represent elements.

The second method for defining a set is implemented by giving a rule or property that a potential element must obey or posses to be included in the set. An example of this is the set

$$A = \{\text{odd numbers between zero and ten}\} \tag{2.2}$$

This is the same set A that was defined explicitly by listing its elements in Eq. (2.1) above. Both of these definitions presuppose the existence of an agreed upon universe of discourse. This collection is called the universal set, it is the collection of objects that are potential members of the sets under discussion. So far we have assumed that the universe of discourse is the natural or counting numbers.

The universe or universal set is usually labeled X or U, although any symbol is allowable. Very often the universal set is not explicitly given as part of the discussion. Instead the universal set is to be inferred from the context of the problem under discussion. If the universe is one common to many fields it may have a standard symbol, such as $\mathbb{N}$ for the natural numbers, $\mathbb{Z}$ for the integers or $\mathbb{R}$ for the real numbers.

The third way to determine a set is through a *characteristic function*. If χ_A is the characteristic function of a set A then χ_A is a function from the universe of discourse X to the set $\{0, 1\}$, where

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{2.3}$$

so that the value 1 indicates membership and the value 0 indicates non–membership.

In the examples above where the set A is the set of odd natural numbers less than ten then the characteristic function for this set is

$$\chi_A(x) = \begin{cases} 1 & x = 1, 3, 5, 7, 9 \\ 0 & \textit{otherwise} \end{cases}$$

The traditional notation for a characteristic function uses the Greek letter χ or *chi* and the set A is indicated as the subscript. However it is desirable for the purposes of this text to introduce a notation that consider A as both the label of a set and as the label of its characteristic function. Thus if A is a set, then its characteristic function is indicated by

$$A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{2.4}$$

The set $\mathbb{B} = \{0, 1\}$ is so important that we give it a special name, the Boole set,

Name	**Symbol**	**Set**
Natural numbers	$\mathbb{N}$	$\{1, 2, 3, \ldots\}$
Bounded natural numbers	$\mathbb{N}_n$	$\{1, 2, 3, \ldots, n-1, n\}$
Non-negative integers	$\mathbb{N}_0$	$\{0, 1, 2, 3, \ldots\}$
Integers	$\mathbb{Z}$	$\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
Rational numbers	$\mathbb{Q}$	$\left\{\frac{a}{b} \mid a, b \in \mathbb{Z}\right\}$
Real numbers	$\mathbb{R}$	any sequence of digits – possibly signed, possibly containing a decimal point, and possibly infinite
Positive real numbers	$\mathbb{R}^+$	$\{x \mid x \in \mathbb{R}$ and $x > 0\}$

Table 2.1.: Common universal sets of numbers in mathematics.

named in honor of Georg Boole who was one of the most important figures in the history of set theory.

A set A is contained in or equal to another set B, written $A \subseteq B$, if every element of A is an element of B, that is, if $x \in A$ implies that $x \in B$. If A is contained in B then A is said to be a *subset* of B and B is said to be a *superset* of A.

Two sets are equal, symbolically $A = B$, if they contain exactly the same elements, therefore if $A \subseteq B$ and $B \subseteq A$ then $A = B$.

If $A \subseteq B$ and A is not equal to B then A is called a proper subset of B, written $A \subset B$. The negation of each of these relations, expressed symbolically by a slash crossing the operator, $x \notin A$, $A \not\subseteq B$, $A \neq B$ and $A \not\subset B$ represent, respectively, x is not an element of A, A is not a subset of B, A is not equal to B and A is not a proper subset of B.

If we are talking about selecting a color for a new tablecloth then the universe is: $X = \{x \mid x \text{ is a color}\}$ read "X equals the set of all elements x such that x is a color" where x is an example element, or variable, that must have the property listed to be contained in the set.

The *intersection* of two sets is a new set that contains every object that is an element of both the set A and the set B.

Example 1. If $A = \{1, 3, 5, 7, 9\}$ and B=$\{1, 2, 3, 4, 5\}$ then the intersection of set A and B is the set $C = A \cap B = \{1, 3, 5\}$, since only the natural numbers 1, 3 and 5 are in both sets A and B.

The *union* of the two sets contains all the elements of either set A or set B.

Example 2. With A and B defined as in Example 1 above $C = A \cup B = \{1, 2, 3, 4, 5, 7, 9\}$. There is no sense in listing an element twice as $\{a\} = \{a, a\}$ since both contain only the single element "a".

The *complement* of a set A, written A^{c}, is the set of all elements of the universe that are not elements of A.

Example 3. Again with A as defined in Example 1 above and with the universal set $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ then the complement of A is $A^{\mathsf{c}} = \{2, 4, 6, 8\}$.

The *relative complement* of a set A relative to B, written $B \setminus A$, is the set of all elements of the set B that are not elements of A. This is also called set subtraction and the notation $B - A$ is often encountered.

Example 4. Again with A and B defined as in Example as defined in Example 1 above then the relative complement of A by B is $B - A = \{7, 9\}$.

The *symmetric difference* of sets A and B, written $A \triangle B$, is the set of all elements that are in only one of the two sets A and B . The notation $A \ominus B$ is also common.

Example 5. Again with A and B defined as in Example as defined in Example 1 above then the symmetric difference of A and B is $A \triangle B = \{2, 4, 7, 9\}$.

The simple framework of set theory allows the basic logical terms of the English language such as "and", "or" and "not" to be translated into precise mathematical expressions. "And" is cast as intersection. "A and B" becomes the intersection of the two sets: $A \cap B$. "Or" is translated as union. "A or B" becomes the union of the set A with set B: $A \cup B$. Lastly the term "not" is rendered as the complement of the set under discussion: "not A" is A^{c}. One last set operation is set difference. $A \setminus B$ is the set of all elements of A that are not elements of B.

All of the concepts of set theory can be recast in terms of the *characteristic functions* of the sets involved. Take for an example the subset relation, where A is a subset of B if and only if the characteristic grade of x in A is less than or equal to the characteristic grade of x in B. In terms of characteristic functions we have that $A \subseteq B$ if and only if $A(X) \leq B(X)$ for all $x \in X$. Remember that we are using $A(x)$ as a shorthand for $\chi_A(x)T$, the characteristic function of A. For the proper subset relation we get strict inequality, $A \subset B$ if and only if $A(x) < B(x)$ for all $x \in X$. The phrase "for all" occurs so often in set theory that a special symbol is used as an abbreviation, $\forall$ represents the phrase "for all". Similarly the phrase "there exists" is abbreviated $\exists$. The definition of set equality is now restated as $A = B$ if and only if

$$\forall x \in X \; A(x) = B(x). \tag{2.5}$$

Intersection can also be defined in terms of characteristic functions. Define the characteristic grade of x in the intersection of two sets A and B *to* be equal to the minimum of the characteristic grade of x in A and of x in B, thus $C = A \cap B$ if and only if

$$\forall x \in X \; C(x) = \min[A(x), B(x)]. \tag{2.6}$$

For the union of two sets A and B, we have $C = A \cup B$ if and only if

$$\forall x \in X \; C(x) = \max[A(x), B(x)]. \tag{2.7}$$

Of course there are different ways to express these relations in terms of characteristic values. The characteristic grade of x in the intersection of A and B is also equal to the product of the characteristic grade of x in A and x in B.

Sets can be finite or infinite. The set of integer numbers is infinite, it goes on without end in both the positive and negative directions. The set of integers is $\mathbb{Z} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$, where "$\cdots$"— the ellipsis, indicates that the values just go on and on towards negative and positive infinity. A finite set is one that contains a finite number of elements. The size of a finite set, called its cardinality, is the number of elements it contains. If $A = \{1, 3, 5, 7, 9\}$ then the cardinality of A, usually written $|A|$, is 5 since A contains five different elements.

A set may contain no elements, the set $\{\}$ contains nothing. This *empty set* is given a special name $\emptyset$, $\emptyset = \{\}$ and $|\emptyset| = 0$. $\emptyset$ is the letter *phi* from the Greek alphabet.

A set can contain another set. Let $D = \{1, 2\}$. The set $E = \{\{1, 2\}, \{1\}, \{2\}, \{\}\}$ contains four elements, it contains the set D as an element. Another element of E is the set $\{1\}$. The empty set, $\emptyset$, is an element of E. Note that $D \in E$ but $D \nsubseteq E$, however both $\emptyset \in E$ and $\emptyset \subseteq E$.

The set of all subsets of a given set X is called the *power set* of X. If X is finite and $|X| = n$ then the number of subsets of X is 2^n. The power set of X is written $\mathcal{P}(X)$ (sometimes the notation 2^X is also used).

Since the next chapter will introduce the concept of a fuzzy set it is often useful to be able to indicate that a set is a classical set and not a fuzzy set. We introduce the term *crisp* to indicate that the set is classical, and specifically, has a characteristic function that maps elements of the universe to the binary set $\{0, 1\}$.

2.3. Intervals

Two special classes of sets are used extensively in the following chapters of this book (and in mathematics as a whole). The first class of sets are the bounded subsets of the real numbers called intervals. The open interval (a, b) contains those real numbers greater than a but less than b so that

$$(a, b) = \{\, x \in \mathbb{R} \mid a < x < b\}$$

The closed interval $[a, b]$ contains those values greater than or equal to a and less than or equal to b,

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$$

The half open (or half closed) intervals are denoted $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$ and

$$[a, b) = \{x \mid x \in \mathbb{R} \mid a \leq x < b\}$$

Of special interest is unit interval $I = [0, 1]$. I contains all the real numbers greater than or equal to zero and less than or equal to one.

The second class of sets are bounded subsets of natural numbers: $\mathbb{N}_n$ will designate the natural numbers less than or equal to some given fixed natural number n. Thus

$$\mathbb{N}_n = \{x \in \mathbb{N} \mid x \leq n\}\,.$$

For example $\mathbb{N}_6 = \{1, 2, 3, 4, 5, 6\}$.

The most import interval for fuzzy set theory is the *unit interval* $\mathbb{I} = [0, 1]$.

2.4. Ordered pairs and relations

Everything in mathematics is defined with set theory as the basis. For example, consider the concept of an ordered pair. The ordered pair is formed from two objects x and y, where $x \in X$ and $y \in Y$, and is denoted $\langle x, y\rangle$. The object x is the first element of the ordered pair and the object y is the second element of the ordered pair.

The *ordered pair* is an excellent example of the use of set theory to create simple definitions for the construction of complex objects. In set theory the ordered pair $\langle x, y\rangle$ is defined to be shorthand for $\{\{x\}, \{x, y\}\}$. The set $\{\{x\}, \{x, y\}\}$ allows us to know which object is the first element in the ordered pair $\langle x, y\rangle$. The first element is the object that is contained in a set of size one in the set definition of an ordered pair. Thus x is the first element of the ordered pair because $|\{x\}| = 1$. The second element of the ordered pair is the object that is contained in the set of size two but is not contained in the set of size one (unless $x = y$). Hence y is the second element of the ordered pair because $y \in \{x, y\}$ but $y \notin \{x\}$. Obviously $\langle x, y\rangle$ is much easier to both write and understand than $\{\{x\}, \{x, y\}\}$. The full set notation is used only to prove that an ordered pair can be defined via set theory and that the ordered pair so defined has all the necessary properties. An example of a property that can be derived from the definition of ordered pair is: $\langle x, y\rangle = \langle y, x\rangle$ if and only if $x = y$.

The set of all ordered pairs where the first element is contained in a set X and the second element is contained in a set Y is called the *Cartesian product* or set product of X and Y and is designated $X \times Y$.

Example 6. If $X = \{1, 2, 3\}$ and $Y = \{a, b\}$ then the Cartesian product of X and Y is $X \times Y = \{\langle 1, a\rangle, \langle 1, b\rangle, \langle 2, a\rangle, \langle 2, b\rangle, \langle 3, a\rangle, \langle 3, b\rangle\}$.

Note that the size of $X \times Y$ is the product of the size of X and the size of Y, $|X \times Y| = |X| \cdot |Y|$. If either set X or Y is empty then the Cartesian product is empty.

Any subset of $X \times Y$ is called a *relation* between X and Y and is designated $r(X, Y)$ or just r. Thus a relation r between X and Y is simply a set of ordered pairs of elements of X and Y. If $\langle x, y\rangle \in r$ then x is said to be related to y or, succinctly, $x\,r\,y$. The in-line notation $x\,r\,y$ is the most popular and readable, but is limited to binary relations, which are the only kind we have introduced so far.

The most common relations in mathematics are the equal relation "$=$" which represents identical objects, and the order relations on the real numbers with the usual representation less than $<$.

Example 7. Let X be the domain of men $\{Tom,\ Dick,\ Harry\}$ and Y the domain of women $\{Eve, Maria, Sally\}$ then the relation "married to" on $X \times Y$ is, for example $\{\langle Tom, Sally\rangle, \langle Dick, Maria\rangle, \langle Harry, Eve\rangle\}$.

A relation from a set X to itself is called a relation on X and $r \in X \times X$.

For more on relations see 8 on page 121.

2.5. Functions

A function is a mathematical abstraction of a consistent machine. Every time you put a specific object into this machine you get an identical product out of the machine.

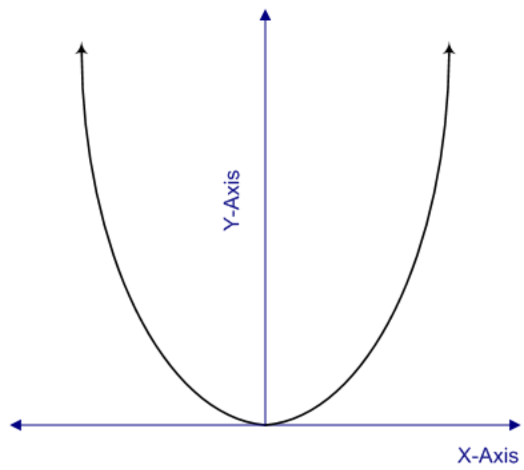


Figure 2.1.: A graph of the function $f(x) = x^2$.

If a relation has a unique second element for each first element then it is called a function. The set X is called the *domain* set and the set Y is called the *range* set. A function is a relation and consequently it is also viewed fundamentally as a set of ordered pairs. The restriction that the function have a unique second element for each first element insures that if $\langle x, y\rangle$ and $\langle x, z\rangle$ are elements of a function (relation) F then $y = z$. This ensures that the function, viewed as a machine, has consistent behavior. A function is said to map a set X into a set Y.

Proposition 1. *Every function is a relation. If $f(x) = y$ then we can also say that $x\,f\,y$*

Definition 1 (domain)**.** The domain of a function f is the set of elements that are mapped, i.e., the set X. Sometimes mathematicians fail to specify the domain and assume the reader can infer the domain set from the context of the discussion.

Definition 2 (co–domain)**.** The co–domain of a function f is the set of elements that can be mapped to, i.e., the set Y.

Definition 3 (image)**.** The image of a function f is the set of elements that are mapped to, i.e., the subset of Y that are the functional values of the elements in X. Sometimes mathematicians use the word range for image.

Usually lower case letters are used for functions. The notation $f : X \to Y$ is used to denote the fact that the function f maps X into Y. Often a mapping rule alone is given to define a function. For example the rule $f(x) = x^2$ is a typical function. Its *implied* domain set is the real numbers and its co–domain is also the real numbers. The set of all values in the co–domain set that are mapped to by elements of the domain set are called the image of the function f or $I(f)$. The image of the function $f(x) = x^2$ is the non-negative reals, symbolically $\mathbb{R}^+$. If $I(f) = Y$ then the function f is called onto. If every ordered pair that defines f is different in both the first and the second element then the function is called one-to-one. A more mathematical way to say this is: if $f(x) = y$ and $f(z) = y$ implies $x = z$ then f is one-to-one. If f is one-to-one and onto then it is possible to define an inverse function f^{-1}:$Y \to X$ such that $f^{-1}(y) = x$ if and only if $f(x) = y$.

It is possible to define the inverse image of a function even if it is not one-to-one and onto. In this case $f^{-1}(y)$ is defined to be the set of elements x that are mapped to y by the function f or $f^{-1}(y) = \{x \mid x \in X \text{ and } f(x) = y\}$.

Example 8. If $X = \{0, 1, 2, 3\}$ and $Y = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ then $f = \{\langle 0, 0\rangle, \langle 1, 1\rangle, \langle 2, 4\rangle, \langle 3, 9\rangle\}$ is a function. Note that the requirements that each $x \in X$ is in some pair in f (f is on X) and that each x is in only one pair in f are both satisfied.

Example 9. Let X be the closed interval $[0, 3]$ and Y the closed interval $[0, 9]$, and $f(x) = x^2$. Then the graph is given in Figure (2.1). A function is continuous if its graph can be drawn with a pencil that never is lifted from the paper. This is not precise mathematics but it gives a feel for what is important.

Other examples of functions are

$$f(x) = x^2 - x - 2, \tag{2.8}$$

$$f(x) = x \bmod 3, \text{ and} \tag{2.9}$$

$$f(x) = \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases}. \tag{2.10}$$

The last function is the absolute value function. It is an example of a function defined for cases. The top row is the case where x is positive or zero, in this case the function is equal to the input; the second row covers the case where x is negative, in this case the function is equal to the negative of the input. Much of mathematics is focused on the study of functions, as almost any college student who has endured calculus knows. This book is also focused on functions, the kind of functions that map a domain set into values that are greater–than or equal to zero and less–than or equal to one. And rest assured that calculus is not needed much at all. There is a quick refresher on the important concepts of calculus in A.1 on page 313.

Suppose that f and g are functions, $f : X \to Y$ and $g : Y \to Z$. Our understanding of functions is that f maps elements of set X to elements of set Y. The function g maps elements of the set Y to elements of the set Z. We can then consider a composite function of f and g that maps elements of X to elements of Z. This function h is usually called the composition of f and g and will be denoted in this book $h = f \circ g$. We can also write that $h(x) = f(g(x))$.

2.6. Distances

One particular useful type of function is used to measure distances. Two examples of distances are city–block and as the crow flies. A city–block distance takes into account that we cannot walk through buildings, in fact, in a city, it is best that we keep to the sidewalks. A crow, or more likely in a city, a pigeon, can fly in a direct line from point A to point B.

A function $dist(A, B)$ that maps two variables A and B, both elements of the same set X, into the non–negative real numbers is called a distance (between A and B), provided that it verifies the following properties.

Positiveness $dist(A, B) \geq 0$ and $dist(A, B) = 0$ if and only if $A = B$.

Symmetry $dist(A, B) = dist(B, A)$.

Triangle Inequality $dist(A, B) + dist(B, C) \geq dist(A, C)$.

Let's consider several examples.

Example 10 (scalar distance)**.** $X = \mathbb{R}$, the set of all real numbers. Then the scalar distance is just

$$dist_H(a, b) = |a - b|,$$

where $|x|$ is the absolute value of x.

Example 11 (Hamming)**.** With $X = \mathbb{R}^2$ the Hamming distance is given by the formula

$$dist(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) = |x_2 - x_1| + |y_2 - y_1|.$$

The Hamming distance is also called the city–block distance. If $X = \mathbb{R}^n$ then the n-dimensional Hamming distance is given by

$$dist_H(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{n} |x_{2i} - x_{1i}|.$$

where $\mathbf{x}$ is an n-dimensional vector in $\mathbb{R}^n$ and n is a natural number. Note that if $n = 1$ then this is the scalar distance.

Example 12 (Euclidean distance)**.** With $X = \mathbb{R}^2$ the Euclidean distance is given by the well known formula

$$dist_E(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

The the Euclidean distance is as the crow flies. If $X = \mathbb{R}^n$ then the n-dimensional Euclidean distance is given by

$$dist_E(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_{i=1}^{n} (x_{2i} - x_{1i})^2}.$$

where $\mathbf{x}$ is an n-dimensional vector in $\mathbb{R}^n$ and n is a natural number.

Example 13 (displacement)**.** With $X = \mathbb{R}^2$ the maximum displacement distance is given by

$$dist_D(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) = max\{|x2 - x1|, |y2 - y1|\}$$

This distance returns the size of the maximum dimension of of an object with corners at $\langle x_1, y_1 \rangle$ and $\langle x_2, y_2 \rangle$.

Example 14 (Minkowski distance)**.** With $X = \mathbb{R}^n$ and p a natural number greater than or equal to two, the Minkowski distance is given by the formula

$$dist_M(\mathbf{x}_1, \mathbf{x}_2) = \sqrt[p]{\sum_{i=1}^{n} (x_{2i} - x_{1i})^p}.$$

where $\mathbf{x}$ is an n-dimensional vector in $\mathbb{R}^n$ and n is a natural number.

2.7. Projections

Suppose we have a subset of a product set, $C \subseteq Z = X \times Y$. Now every subset of a product set is considered a relation. It is useful to have a notation for the projections into the first and second dimension. The projection of set C into the first dimension is the set of all $x \in X$ such that x occurs as the first element of some ordered pair $\langle x, y \rangle$ that is contained in C;

$$proj_1(C) = \{x \mid x \in X \text{ and } \exists\, y \in Y \text{ with } \langle x, y \rangle \in C\} \ . \tag{2.11}$$

The projection into Y, or into the second dimension is

$$proj_2(C) = \{y \mid y \in Y \text{ and } \exists\, x \in X \text{ with } \langle x, y \rangle \in C\} \ . \tag{2.12}$$

If $A \subseteq X$ is a set then its cylindric extension to Z is the set $A \times Y$. If $B \subseteq Y$ is a set then its cylindric extension to Z is the set $X \times B$. If $A \subseteq X$ and $B \subseteq Y$ are both sets in different dimensions (different universal sets) then their cylindric closure is the intersection of their cylindric extensions. It is not hard to show that the cylindric closure of A and B is $A \times B$.

We can attempt to reconstruct $C \subseteq X \times Y$ from its projections by constructing its cylindric closure but it is not hard to see that this set is not the original relation, but that cylindric closure is larger. In effect, the cylindric closure fills in the “missing” pieces.

$$C \subseteq proj_1(C) \times proj_2(C) \ . \tag{2.13}$$

Example 15. Let $X = \{1, 2, 3\}$, $Y = \{a, b, c\}$ and $C = \{\langle 1, a \rangle, \langle 2, b \rangle\}$. Then

$$proj_1(C) = \{1, 2\} \ . \tag{2.14}$$

and

$$proj_2(C) = \{a, b\} \ . \tag{2.15}$$

Finally comparing C to its cylindric closure shows that

$$\{\langle 1, a \rangle, \langle 2, b \rangle\} = C \subseteq proj_1(C) \times proj_2(C) = \{\langle 1, a \rangle, \langle 2, a \rangle, \langle 1, b \rangle, \langle 2, b \rangle\} \tag{2.16}$$

.

2.8. Extension

Let X and Y be non-empty universal sets. Let $f : X \to Y$ be a function from X into Y. Let r be a relation between elements of X and elements of Y, that is $r \subseteq X \times Y$.

Let $G \subseteq X$ and $H \subseteq Y$. The standard extensions of the function f and relation r to set arguments are

$$f(G) = \{y \mid x \in G \text{ and } y \in Y \text{ and } f(x) = y\} \tag{2.17}$$

and

$$r(G) = \{y \mid x \in G \text{ and } y \in Y \text{ and } x\, r\, y\} \tag{2.18}$$

It is traditional to use the same symbol for the function (relation) and its set extension since the argument indicates which case applies. Note that these extensions are now functions and relations on power sets of X and Y. That is $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ and $r \subseteq \mathcal{P}(X) \times \mathcal{P}(Y)$. We can also define formal inverses for f and r

$$f^{-1}(y) = \{x \mid x \in X \text{ and } f(x) = y\} \tag{2.19}$$

and

$$r^{-1}(y) = \{x \mid x \in X \text{ and } x\, r\, y\} \tag{2.20}$$

that are themselves naturally set valued. The extension of the function f and relation r to set arguments are

$$f^{-1}(H) = \{x \mid x \in X \text{ and } y \in H \text{ and } f(x) = y\} \tag{2.21}$$

and

$$r^{-1}(H) = \{x \mid x \in X \text{ and } y \in H \text{ and } x\, r\, y\} \tag{2.22}$$

where $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$ and $r^{-1} \subseteq \mathcal{P}(Y) \times \mathcal{P}(X)$.

In the final analysis, we only needed to introduce the formulas for the extension of relations, since functions are really special cases of relations. However, in mathematics, functions often play a larger role than relations, and so the extension of a function is made explicit in formula 2.17.

2.9. Homework

Let us define the universal sets

$$X = \{1, 2, 3, 4, 5, 6\}, \tag{2.23}$$

$$Y = \{a, b, c\}, \tag{2.24}$$

and

$$Z = \{\alpha, \beta, \gamma, \delta\}. \tag{2.25}$$

Let

$$\begin{aligned} A &= \{1,2\}, \\ B &= \{1,3\}, \\ C &= \{2,4,6\}, \\ D &= \{a,b\}, \\ S &= \{\langle 1,a\rangle, \langle 2,b\rangle, \langle 3,b\rangle, \langle 2,c\rangle\} \text{ and} \\ T &= \{\langle a,\delta\rangle, \langle b,\gamma\rangle, \langle c,\alpha\rangle\}. \end{aligned}$$

Given the above information, answer the following questions.

1. For the sets, A through T, what universe do they belong to?
2. What is A^{c} ?
3. What is B^{c} ?
4. What is C^{c} ?
5. What is $A \cup B$?
6. What is $A \cup C$?
7. What is $B \cup C$?
8. What is $A \cup B^{\mathsf{c}}$?
9. What is $A \cup C^{\mathsf{c}}$?
10. What is $B \cup C^{\mathsf{c}}$?
11. What is $A \cup A^{\mathsf{c}}$?
12. What is $A \cap B$?
13. What is $A \cap C$?
14. What is $B \cap C$?
15. What is $A \cap B^{\mathsf{c}}$?
16. What is $A \cap C^{\mathsf{c}}$?
17. What is $B \cap C^{\mathsf{c}}$?
18. What is $A \cap A^{\mathsf{c}}$?
19. What is $A \setminus B$ and $B \setminus A$.
20. What is $A \setminus C$ and $C \setminus A$.
21. What is $B \setminus C$ and $C \setminus B$.
22. What is $A \triangle B$.

23. What is $A \triangle C$.
24. What is $B \triangle C$.
25. What is $A \times B$?
26. What is $A \times C$?
27. What is $A \times D$?
28. What is $D \times A$?
29. What is $B \times C$?
30. What is $S \circ T$?
31. What is the projection of T into Y ?
32. What is the projection of T into Z ?
33. What is the cross product of the projections in (31) and (32) ?

3. Logic

3.1. Introduction

In ancient Greece, before the age of radios, TV, or iPods, entertainment consisted of going down to the Acropolis, drinking wine, eating bread and cheese, and arguing. About 350 BC, Aristotle noticed something interesting about these conversations. Aristotle noticed that the pattern of an argument did not depend on the subject of the argument.

For example, let us compare the two following arguments:

It is raining.	*Premise 1*
If it rains the grass gets wet.	*Premise 2*
The grass is wet.	*Conclusion*

and

Aristotle is a man.	*Premise 1*
All men are mortal.	*Premise 2*
Aristotle is mortal.	*Conclusion*

Here is the pattern Aristotle noticed, abstracted from the particular subject, and using letters A and B to represent propositions, just like Aristotle did in his writings.

A	*Premise 1*
if A *then* B	*Premise 2*
B	*Conclusion*

Aristotle called a logical deduction pattern a syllogism. We called this particular pattern the modus ponens (Latin for *the way that affirms*).

The most important thing about logic is that it is designed to insure that when the premises are true then the conclusion that is deduced must be true. If any of the premises are false logic can not tell you anything about the truth or falsity of the conclusions. Thus a logical conclusion eliminates uncertainty about the reasoning process. You always have to be careful with logic though:

A ham sandwich is better than nothing.	*Premise 1*
Nothing is better than complete happiness.	*Premise 2*
A ham sandwich is better than complete happiness.	*Conclusion*

Here, the problem is that in English, the word nothing has two different meanings in the two Premises.

	A	B	$A \to B$
Row 1	0	0	1
Row 2	0	1	1
Row 3	1	0	0
Row 4	1	1	1

Table 3.1.: The truth table for the logical connective: implication.

3.2. Symbolic logic

Classical logic is a system designed to ensure that arguments reach correct conclusions. A logical system starts with a set of axiomatic assumptions, such as "two points determine a line." Axioms are assumed to be true. Logic then produces theorems which are guaranteed to be true as long as the axioms are true and do not contradict themselves.

The problems with classical logic are mainly twofold. The first is determining what should and should not be axioms, that is, what can we assume to be true. The second problem is that logic, like mathematics, is abstract, and does not fully mirror the real world. When I say "I will meet you at 1PM tomorrow" is this statement true or false?

Classical logic as a system was invented by Aristotle and his works on this subject are collected in a book called the *Organum*. Aristotle uses letters of the alphabet to stand for primitive objects that can be either true or false. Thus A might stand for the fact that the apple is red. Aristotle presents rules for reasoning about these objects, i.e., how to come to conclusions that are entailed by the axioms.

Modern symbolic logic originated at the turn of the 19th century when symbols were introduced for the logical expressions. For example, the statement "If the apple is red and the sky is blue then have a picnic" becomes $A \wedge B \to C$ where "$\wedge$" represents logical conjunction (and) and "$\to$" represents implication (if – then). Two types of logic are of primary interest to us in this chapter. They are first and second order logics, also called propositional and predicate logics. First order logic deals with Propositions that are either true or false, such as "The apple is red." Second order logic deals with Predicates such as "The object is red," which is true or false depending on what object we are talking about.

3.2.1. First order logic

Symbolic logic uses a set of symbols to produce expressions (called well formed formulas or *wff*s) that have two possible values when evaluated. These values are called true or false and are commonly represented mathematically with the numeric values 1 and 0. The symbols used include:

Symbol	English
$\neg$	not
$\wedge$	and
$\vee$	or
$\rightarrow$	implies

3.2.2. Truth tables

We will focus for a while on what we mean by a logical conjunction or and with its symbolic representation $\wedge$. In English we would write "The sky is blue and the apple is red." In logic this would become $A \wedge B$. When we look at the sky it may be blue or it may be grey. When we substitute A for this statement we have a propositional variable which may be true or false. Similarly B represents the statement that the apple is red which can also be true or false. In logic, the rule for and is; the and of A and B is true only if both A and B are simultaneously true, otherwise the and of A and B is false. The traditional method of presenting this rule was as a truth table:

	A	B	$A \wedge B$
Row 1	true	true	true
Row 2	true	false	false
Row 3	false	true	false
Row 4	false	false	false

Row 1 in this table is interpreted as saying if A is true and B is true then the conjunction $A \wedge B$ is true. Rows 2–3 capture the other three possibilities, where at least one of A or B is false.

In the modern computer world, things are presented slightly differently. First off, computers use numbers to represent everything (the letter A is represented by 65), so we represent false with zero and truth with one. Since zero is less than one zero comes first. Therefore the tables we will use, remembering that true $\equiv 1$ and false $\equiv 0$, will look like:

	A	B	$A \wedge B$
Row 1	0	0	0
Row 2	0	1	0
Row 3	1	0	0
Row 4	1	1	1

This table presents the same information, the rules for conjunction, in a slightly different form.

The rule for disjunction, or or, is that $A \vee B$ is true if either side is true. Negation, which operates on a single proposition, toggles the value from true to false and back again. Implication has a truth table that needs some explanation.

3.2.2.1. Implication

Implication, if A then B, is modeled in classical logic with the arrow operator $A \to B$. The result of implication for every possible truth assignment is given in the Truth Table (3.1).

Let us give a justification of the Table (3.1) due to Łukasiewicz. Consider the statement about the natural numbers

$$\text{if } n \text{ is divisible by } 6 \text{ then } n \text{ is divisible by } 3\ . \tag{3.1}$$

To any reasonable person this is a universally true statement about the natural numbers (the counting numbers, $1, 2, 3, \dots$). Let us examine it for a few numbers n. If n is 12 then the truth value of n is divisible by 6 is true as is the truth vale of n is divisible by 3, so if truth then truth produces truth (row 4 of Table (3.1)). If n is 13 then the truth value of n is divisible by 6 is false as is the truth vale of n is divisible by 3, so if false then false produces truth (row 1 of Table (3.1)). If n is 15 then the truth value of n is divisible by 6 is false but the truth vale of n is divisible by 3 is truth, so if false then truth produces truth (row 2 of Table (3.1)). However it is impossible to produce a number n that will generate row 3 of Table (3.1). That is because the statement is a universal truth.

The rules for all the most common connectives are presented in Table 3.2.

3.2.2.2. Compound sentences

Now we will consider a more complex logical expression:

$$\neg B \vee A \to A \wedge B \tag{3.2}$$

Using Table 3.2 we can calculate the resultant truth value of a statement like Eq. (3.2) for arbitrary true $= 1$ or false $= 0$ valuations of its atomic variables A, B, C, $\dots$. In mathematics variables can represent complex expressions, The table applies to anything of the form $W \to Q$ event though the table uses A and B. The table says that if we have something of the form $W \to Q$ and that Q is false while W is true then we can conclude that the whole statement $W \to Q$ is false. But W might represent $\neg B \vee A$ and Q might represent $B \wedge A$. In Eq. (3.2) the atomic variables A, B, and C represent statements that can be true or false.

To calculate the truth values for more complicated statements, such as that in Eq. 3.2 it is best to use a table format. In the table below we start by adding columns for all the atomic variables in the statement. In this case the atomic variables are A and B so we put these in the first two columns (second row), and label those columns a and b (first row). We then write out the logical equation $\neg B \vee A \to A \wedge B$ giving each individual symbol its own column. (second row, columns 3–13). Next we list all the possible combinations of true and false that the atomic variables can take, so that the first two columns gain four rows. Each atomic variable can have only two values so that there will be 2^nrows when there are n atomic variables.

This is the initial setup. We now fill in the columns labeled $a1$, $a2$, $b1$ and $b2$ that represent the values of the atomic variables. Column $b1$ takes the values of B from column b. Similarly column $b2$ takes the identical values of B copied from column b. Columns $a1$ and $a2$ take the values of A from the column labeled a. The rules for logical connectives say that negation is done first, then conjunction, then disjunction, and finally implication, symbolically “$\neg$”, then “$\wedge$”, then “$\vee$”’, and finally “$\rightarrow$”. Of course, expressions in parentheses must be done first (just like in arithmetic). Thus the first calculation is done in the column labeled 1 where we calculate $\neg B$ based on the values of B provided by column $b1$. The second calculation produces column 2 which evaluates $A \wedge B$ using the values in columns $a2$ and $b2$. The third calculation produces column 3 which evaluates $\neg B \vee A$ using the values in columns 1 and $a1$. Finally the fourth calculation produces column 4 which evaluates $\neg B \vee A \rightarrow A \wedge B$ using the values in columns 3 and 2. When doing this on paper, it is easy to cross out each column as you use it. Thus after the first step you can cross out column $b1$.

The final result (column 4) is that $\neg B \vee A \rightarrow A \wedge B$ is sometimes true, as sown in rows 3, 5, and 7, and sometimes false, as in row 4.

a	b		1	$b1$	3	$a1$		4		$a2$	2	$b2$	
A	B	(	$\neg$	B	$\vee$	A	)	$\rightarrow$	(	A	$\wedge$	B	)
0	0		1	0	1	0		1		0	0	0	
0	1		0	1	0	0		0		0	0	1	
1	0		1	0	1	1		1		1	0	0	
1	1		0	1	1	1		1		1	1	1	

If the formula is true for all values of A, B, C then the formula is called a tautology. For example

$$\neg A \wedge \neg B = \neg(A \vee B)$$

is a famous tautology called the De Morgan law. The other De Morgan law is

$$\neg A \vee \neg B = \neg(A \wedge B)\ .$$

Two other important tautologies are the law of the excluded middle

$$A \vee \neg A$$

and the law of contradiction

$$\neg\,(A \wedge \neg A)\ , \tag{3.3}$$

If a statement is always true, a tautology, then it is a new truth, and can be used as such in derivations. Note that above we have proved that $\neg B \vee A \rightarrow A \wedge B$ is not a tautology, hence it is not a truth.

Logic is one of those subjects that just invites argument in every sense of the word. Aristotle said “A truth can only be derived from previously known truths.” The problem is, where do we start? How do we get first truths? In mathematics these known truths are called axioms. They are reasonable statements that can be accepted without justification. Except, every axiom of every formal system has been questioned in numerous texts. And some axioms, such as the axiom of choice, that one can randomly choose an example element from a set, which is absolutely indispensable to most of mathematics, leads to conclusions that are not easily accepted (the well

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$	$A = B$	$A \oplus B$
0	0	1	0	0	1	1	0
0	1	1	0	1	1	0	1
1	0	0	0	1	0	0	1
1	1	0	1	1	1	1	0

Table 3.2.: The calculation rules for logical connectives

ordering theorem). "Where logic deals with ideals and abstractions it can have no meaning" This is a direct quote from one of the most famous logicians who ever lived, Bertram Russel. Thus classical logic is abstract and its results are meaningless – **literally meaningless.**

Numbers for example have no physical representation. There can be two apples but there is never a physical two. An apple is a thing you can see and feel and eat, a two cannot be seen, felt, tasted, etc. it can only be imagined. The variable v can be two but can v be an apple? No, a map is not a country, a picture is not a mountain, a poem is not a tree. There are even worse problems with logic, Godël *proved* that you could not prove all the truths that a formal logic could express, unless the logic contained paradox, a statement the was both true and false like "Everything I say is a lie".

3.3. Predicate logic

Equation 3.1 is actually an example of predicate logic.

Predicate logic deals with predicates $A(x)$ that make a statement (a predicate) about x having the property A. The object x is limited to the Universe of Discourse, usually some set. In Equation 3.1 we are assuming the universe is the Natural Numbers and that we have two predicates, one says x is divisible by three, $T(x)$, and the other says that x is divisible by six, $S(x)$. The symbolic representation of Equation 3.1 is $S(x) \rightarrow T(x)$.

"Socrates is a man" means that the statement that object $x \equiv socrates$ possesses the property $A \equiv manness$ is true. Predicate logic, also called second order logic, sees all the symbols of first order logic, with a few additions.

Symbol	English
$\neg$	not
$\wedge$	and
$\vee$	or
$\rightarrow$	implies
$\forall$	for all
$\exists$	there exists

In predicate logic there are two extra symbols of great importance called quantifiers. The symbol $\exists$ stands for "there exists" and the symbol $\forall$ is read as "for all".

$\forall x\,(x \geq 0)$ is read that "for all x, x is greater than or equal to zero." If the universe of discussion is the integers $\mathbb{Z}$ then this statement is false. If the universe of discussion is the natural numbers$\mathbb{N}$ then this statement is true.

$\exists x\,(x \geq 0)$ is read there exists an x, such that x is greater than or equal to zero. This statement is true if the universe of discussion is the natural numbers $\mathbb{N}$ or the integers $\mathbb{Z}$.

In predicate logic all predicates $A(x)$, $B(x)$ are either true or false for a particular x in the universe of discourse.

Definition 4 (universal quantifier)**.** $\forall x\, A(x)$ is true if $A(x)$ is true for every x in the universe of discourse.

Definition 5 (existential quantifier)**.** $\exists x\, A(x)$ is true if $A(x)$ is true for any x in the universe of discourse.

Thereafter everything proceeds as in first order logic using the standard logical connectives. Much more can be said about logic but this is sufficient for our purposes. Fuzzy logic awaits us in Chapter 15.

3.4. Derivations

The first, and most important, deduction scheme is the modus ponens which is written in tableau form as;

$$\begin{array}{l} A \\ A \rightarrow B \\ \hline B \end{array} \tag{3.4}$$

and whose meaning is exemplified in Table 3.3a.

This deduction is expressed in symbolic logic by the formula

$$(A \wedge (A \rightarrow B)) \rightarrow B\;. \tag{3.5}$$

Second in our list of deduction schemes is the modus tolenswhich is written in tableau form as;

$$\begin{array}{l} \neg B \\ A \rightarrow B \\ \hline \neg A \end{array} \tag{3.6}$$

A	Premise 1
if A then B	Premise 2
B	Conclusion

(a) modus ponens

not B	Premise 1
if A then B	Premise 2
not A	Conclusion

(b) modus tolens

if A then B	Premise 1
if B then C	Premise 2
if A then C	Conclusion

(c) hypothetical syllogism

Table 3.3.: Rules of derivation.

and whose meaning is exemplified in Table 3.3b.

This deduction is expressed in symbolic logic by the formula

$$\neg B \wedge (A \to B) \to \neg A \ . \tag{3.7}$$

The third deduction scheme of interest is the hypothetical syllogism, which is written in tableau form as;

$$\begin{array}{c} A \to B \\ B \to C \\ \hline A \to C \end{array} \tag{3.8}$$

and whose meaning is exemplified in Table 3.3c.

This deduction is expressed in symbolic logic by the formula

$$((A \to B) \wedge (B \to C)) \to (A \to C) \ . \tag{3.9}$$

The most important part of classical logic was the ability to derive new truths from old truths.. A method of derivation is called a deduction rule. The most famous is the modus ponens which is sometimes written

$$A \wedge (A \to B) \to B$$

but this form ignores the fact that we have two distinct facts. One fact is that A is true. The other fact is that the truth of A implies the truth of B. From these two facts we arrive at the conclusion that B is true. Thus the rule is more properly is written

$$\begin{array}{r} A \\ \underline{A \to B} \\ B \end{array}$$

and says that if A is true and A implies B is true then we can logically conclude that

B is true. We can illustrate this derivation thusly:

It is raining.
If it rains the grass gets wet.
The grass is wet.

The Latin names for these rules of deduction represent the fact that logic was one of the few areas of knowledge kept alive during the dark ages of western Europe. The logical method called the modus tollens is properly written as:

$$\begin{array}{c} A \rightarrow B \\ \underline{B \rightarrow C} \\ A \rightarrow C \end{array}$$

This line of reasoning is usually illustrated with the example:

Socrates is a man.
All men are mortal.
Socrates is a mortal.

3.5. Logic and set theory

We are going to examine this correspondence in detail, since it is very important to the development of fuzzy set theory and fuzzy logic.

The basic correspondence we wish to exploit is that between the characteristic function of a subset of a universe and a predicate.

Let the universe of discourse be $U = \{a, b, c, d, e, f\}$. Let $A = \{a, e\}$ be the set of vowels in the universe and $B = \{d, e, f\}$ the set of the last three letters. It is easy to see that $A \cap B = \{e\}$. Remember that the characteristic function of a set is zero for x if x is not in the set and is one if x is in the set. Let us examine the relation between A, B, and $A \cap B$ in terms of their characteristic functions. Consider the element a of U. Since it is in A we have that $\chi_A(a) = 1$, since it is not in b we have that $\chi_B(a) = 0$, and finally, since a is not in the intersection of A and B, we have that $\chi_{A \cap B}(a) = 0$. This pattern will always hold, that is, if $x \in U$ then if characteristic function for the first set is one and the second set is zero, , $\chi_A(x) = 1$ and $\chi_B(x) = 0$, then we will always have that the element is not in the intersection and that the characteristic function of the intersection is zero for that x, that is $\chi_{A \cap B}(x) = 0$. Table 3.4a shows the four possible patterns for $\chi_A(x)$ and $\chi_B(x)$ as x cycles through U. When $x = a$ we have Case 3 as illustrated previously. Elements b and c are covered by Case 1. Elements d and f are covered by Case 2. Finally elements e is covered by Case 4.

Let the universe of discourse be $U = \{a, b, c, d, e, f\}$. Let A be the predicate that $x \in U$ is a vowel in the universe and B the predicate that $x \in U$ is one of the last three letters in the universe. It is easy to see that $A \cap B = \{e\}$. Now if a predicate A is true $\equiv 1$ about x and predicate B is false $\equiv 0$ about x then the conjunction operator produces a false $\equiv 0$ for their and. In other words true and false equals false. Table 3.4b shows the four possible patterns for $A(x)$ and $B(x)$ as x cycles through U. When $x = a$ we have Case 3 as illustrated previously. Elements b and c are covered by Case 1. Elements d and f are covered by Case 2. Finally elements e is covered by Case 4.

It seems that the behavior of set intersection and logical and are very similar. Mathematics calls this an isomorphism, which is Greek for same shape. It turns out that

Case	$\chi_A(x)$	$\chi_B(x)$	$\chi_{A\cap B}(x)$
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	1

Case	$A(x)$	$B(x)$	$A \wedge B(x)$
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	1

Table 3.4.: Tables for the intersection operator and logical and.

Logical connective	English translation	Set operation
$\sim A$ or $\neg A$	not A	A or $\overline{A}$
$A \wedge B$	A and B	$A \cap B$
$A \vee B$	A or B	$A \cup B$
$A \rightarrow B$	if A then B	$A \subseteq B$
$A = B$	A equals B	$A = B$
$A \oplus B$	A exclusive or B	$A \bigtriangleup B$

Table 3.5.: Logical connectives, with English translations and set operation equivalences.

set complement and logical negation are also isomorphic. So is set union and logical or.

In a slightly different way subsethood mimics logical implication. The difference is that subsethood is a relation while implication is an operator. Still if $A(x) \rightarrow B(x)$ is a tautology, that is it is always true, then the set of elements that makes A true will be a subset of the set of elements that make B true.

Table (3.5) summarizes the isomorphism between set theory and logic. It gives the English meaning and the corresponding set theory operators and logical connectives.

The purpose of this chapter is to extend classical logic to fuzzy logic as classical or crisp sets were extended to fuzzy sets. However, as will be seen we may have to abandon the truth table () in its entirety. It is also important to note that this is not the first attempt to extend classical logic. Many others have tried, including Lukasiewicz who developed a three valued logic for true, false, and unknown as well as multi-valued logics. The logic of two-values is often called Boolean logic.

3.5.1. Other logics

The limitations of formal logic led to many attempts to expand their role and power. Some of the attempts include:

DEONTIC LOGIC —Logics of permission and obligation (derived from modal logics of possibility and necessity); hence the logic of norms and normative systems.

EPISTEMIC LOGIC —The logic of non-truth-functional operators such as “believes” and “knows”. For example, let $\ddot{p}$ mean that “I know proposition p”. If $\ddot{p}$ and $p \rightarrow q$ are given, then what must we add in order to infer $\ddot{q}$?

INTUITIONISTIC LOGIC —Propositional logics (and their predicate logic extensions) in which neither "$p \vee \tilde{}p$" nor "$\neg\neg p \rightarrow p$" are provable. They accept disjunctions $A \vee B$ as theorems only if one of the disjuncts is separately provable: i.e. if either $\vdash A$ or $\vdash B$. They have the same rules of inference as classical logic. Propositional connectives are undefined primitives.

MULTIVALUED LOGICS —Logics in which there are more than the two standard truth-values "truth" and "falsehood". Motivated by semantic paradoxes like the liar ("this statement is false") and by future contingents ("tomorrow there will be a sea-battle"), that don't easily take either standard truth-value, and by attempts to deal with uncertainty, ignorance, and "fuzziness". An early example was Lukasiewicz's three valued logic with atomic variables that are true, false, or undetermined (since the original is in Polish, the third value has many other translations, such as possible, and unknown). Thus the image set of the characteristic function might be the set $T = \{0, u, 1\}$, and u is usually mapped to the numerical value $\frac{1}{2}$.

MODAL LOGIC –necessary and possibly true. There are two additional unary operators $\Box p$ and $\Diamond p$ that represent a necessarily and possibly true proposition. This is based upon an all possible worlds framework.

QUANTUM LOGIC —To reflect quantum indeterminacy and uncertainty, quantum logic adds a third truth-value ("indeterminate"); hence the metatheory denies the principle of excluded middle (PEM). Nevertheless, for every p, "$p \vee \neg p$" is logically valid in systems of quantum logic. That is, PEM is true in the theory, false in the metatheory. Because both disjuncts of a true disjunction can be false, disjunction and conjunction behave asymmetrically; hence the distribution laws generally fail. Motivated to capture the queerness of quantum-mechanics; in quantum logic this queerness shows up on the propositional level, in redefined connectives.

TEMPORAL (TENSE) LOGIC —Time dependant truth. Logics in which the times at which propositions bear certain truth-values can be indicated, in which the "tense" of the assertion can be indicated, and in which truth-values can be affected by the passage of time.

3.6. Multi-valued logics

Multi-valued logics are the precursor to fuzzy logic. Łukasiewicz was the first to introduce a three valued logic containing true, false, and unknown (or undecided or undetermined). Typically, these truth values are mapped to the numerical values true $\equiv 1$, false $\equiv 0$, and unknown $\equiv \frac{1}{2}$ (or undetermined). The problem in designing multi-valued logics is defining the logical connectives, especially the implication operator "$\rightarrow$". For a three valued logic there are many large truth tables to memorize. What if we have a five valued logic, to model a Likert scale with $\mathsf{strong-disagree} \equiv 0$, $\mathsf{disagree} \equiv \frac{1}{4}$, $\mathsf{neutral} \equiv \frac{1}{2}$, $\mathsf{agree} \equiv \frac{3}{4}$, and $\mathsf{strong-agree} \equiv 1$. Truth tables for this system have 25 lines to them.

Tables don't even work for a function from atomic variables that have continuous truth values. Tables don't work for a domain where a predicate $A(x)$ can have any truth value between false $\equiv 0$ and true $\equiv 1$, which is exactly what fuzzy logic needs and uses.

However, sometimes the truth tables of these various logics can be avoided by the use of formulas.

Example 16 (Lukasiewicz 3–valued logical connectives)**.** Let t represent true, u represent undetermined, and f represent false. The truth table for the negation operator for Łukasiewicz 3-valued logic is

$\neg$	
t	f
u	u
f	t

.

The truth table for the implication operator, $A \to B$, in Łukasiewicz 3-valued logic is

$\to$	t	u	f
t	t	u	f
u	t	t	u
f	t	t	t

(3.10)

where the A value comes from the row heading and the B value comes from the column heading.

Example 17. If we associate the numerical values of $t = 1$, $u = \frac{1}{2}$, and $f = 0$ in Łukasiewicz 3-valued logic then

$$p \to q = 4p^2q^2 - 4p^2q - 4pq^2 + 5pq - p + 1$$

whenever $p, q \in \left\{0, \frac{1}{2}, 1\right\}$.

Example 18. If we associate the numerical values of $t = 1$, $u = \frac{1}{2}$, and $f = 0$ in Łukasiewicz 3-valued logic then

$$p \to q = \min[1, 1 - p + q]$$

whenever $p, q \in \left\{0, \frac{1}{2}, 1\right\}$.

In the last two examples, the polynomial in two variables $4p^2q^2 - 4p^2q - 4pq^2 + 5pq - p + 1$ provides an analytic expression for Łukasiewicz implication that is a continuous, and differentiable; on the other hand, $\min[1, 1 - p + q]$ is much simpler, however the $\min$ function is continuous but not differentiable.

Example 19. If we associate the numerical values of $t = 1$, $u = \frac{1}{2}$, and $f = 0$ in Łukasiewicz 3-valued logic then:

$$\begin{aligned} p \wedge q &= \min[p, q] \\ p \vee q &= \max[0, p + q - 1] \\ \neg p &= 1 - p \end{aligned}$$

whenever $p, q \in \left\{0, \frac{1}{2}, 1\right\}$.

Example 20 (Godel 3–valued logical connectives)**.** Let t represent true, u represent undetermined, and f represent false. The truth table for the negation operator for Godel 3-valued logic is

$\neg$	
t	f
u	f
f	t

.

The truth table for the implication operator, $A \to B$, in Godel three-valued logic is

$\to$	t	u	f
t	t	u	f
u	t	t	f
f	t	t	t

(3.11)

where the A value comes from the row heading and the B value comes from the column heading. In addition

$$p \wedge q = \min[p, q]$$
$$p \vee q = \max[p, q]$$

3.7. Notes

The demonstration for the truth table of implication comes from Łukasiewicz (1963) who also introduced three valued Łukasiewicz (1920). Multi-valued logics were introduced by Post (1921).

3.8. Homework

1. Seek out a fellow student that is taking logic in the philosophy departement at your college. Ask them what notation their book uses for the implication operator. Ask them what the course is supposed to teach them?
2. Seek out a fellow student that is taking logic in the math or computer science departement at your college. Ask them what notation their book uses for the implication operator. Ask them what the course is supposed to teach them?
3. Go to www.amazon.com and do a book search for logic. Why this modern obsession with logic. Did you find any logics not described in this chapter?
4. Is $A \to (B \to A)$ a tautology?
5. Is $A \to (A \to B)$ a tautology?

6. Is $(\neg A \vee B) = (A \rightarrow B)$ a tautology?
7. What is wrong with the ham sandwich example?
8. Since we can prove that the world is round why do some people still beleive it is flat.
9. How many angels can dance on the head of a pin. Prove your answer.

4. Probability Theory

4.1. Introduction

> I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all. Ecclesiastes 9:11

That chance plays a large part in our lives has been known for a long time. In the 18th century the origins of probability theory began when French mathematicians such as Pascal and Fermat tried to provide precise answer to questions about games of chance.

Probability theory deals with random processes. A random process is one whose exact outcome in a single experiment is indeterminate, but whose long term behavior in repeated experiments is describable. The classic example is the outcome of flipping a coin. It is impossible to know whether a fair coin thrown into the air will fall to the ground exposing the side called heads or the side called tails. However if we flip the same coin one thousand times we expect that both the number of heads and the number of tails will be very close to five hundred each. This is a description of long term behavior; in repeated coin flips we expect the frequency of the heads and the frequency of the tails to be almost equal. The flipping of a coin is therefore considered a random event.

Bayesian Probability deals with subjective belief. Consider going to Las vegas and betting on Detroit Lions to win Superbowl XLIII. before the season starts. The Las Vegas oddsmakers consider the chance of the Lions winning the Superbowl as remote, and sets the odds at 100/1 (one–hundred to one). Odds of 100/1 means a bet of $1 would return $100 if Detroit does in fact win the Superbowl. The Detroit Lions winning Superbowl XLIII is not a repeatable experiment. These odds are set based on the knowledge and experience of the oddsmaker. As soon as the NFL season starts, games are won and lost, and these facts change the odds. If Detroit starts winning the odds narrow and a better gets a smaller return for a $1 bet.

Probability

Here we present Probability as it is used for dealing with random processes.

Probaility theory comes with its own special jargon. The universal set s is called a sample space, its elements are called sample outcomes and its subsets are called events. If A and B are subset of the sample space (events) then the event "A or B" is modeled as $A \cup B$, the event "A and B" happened is modeled as $A \cap B$, and the event "not A" happened is modeled as the complement of A . While the first two models for "or" and "and" use standard set notation for unions and intersections, the third, "not", borrows from the notation of logic, and most probability books use either $\bar{A}$ or A' for negation. Finally, probability uses a notation that is essentially undefined! The probability of event A happening given that B is already known to have occured is denoted $P(A \mid B)$, which is read as the probability of A given B. The probability $P(A \mid B)$ is called a conditional probability.

Given a finite sample space S with $|S| = n$ and a function p that maps this set into the unit interval,

$$p : X \to I$$

then p is called a finite probability distribution iff

$$\sum_{s \in S} p(s) = 1$$

A finite probability measure is a mapping from the set of all events E (subsets of S) into the unit interval,

$$P : E \to I$$

such that the probability measure of the empty set is zero:

$$P(\emptyset) = 0 \tag{4.1}$$

the probability measure of the sample space is one:

$$P(S) = 1 \tag{4.2}$$

and the measure is additive, that is, the probability of A or B (modeled as $A \cup B$) is equal to the sum of the probailities of A and B minus the probability of A and B (modeled as $A \cap B$):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{4.3}$$

This last rule is easy to see from a Venn Diagram of A and B. If we think of the area of the circles A and B as their probability, then the are of $A \cup B$ is the area of A plus the area of B minus the overlap $A \cap B$, which had been counted twice by adding the probabilities of A and B. See figure (4.1).

There is a one-to-one correspondence between finite probability distributions and finite probability measures. Given a finite probability measure, P, we can derive the associated finite probability distribution by way of the formula

$$p(s) = P(\{s\}).$$

In the other direction, the probability assigned to a subset of the universal set is equal to the sum of the probabilities that the distribution assigns to the elements of that subset,

$$P(A) = \sum_{s \in A} p(s).$$

For simplicity of notation let us define

$$p_i = p(s_i) = P(\{s_i\})$$

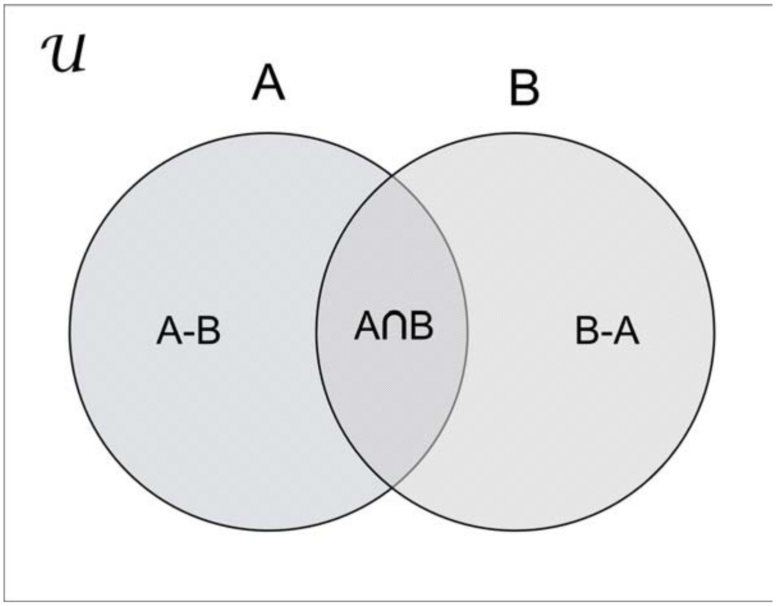


Figure 4.1.: Venn digram of $A \cup B$.

for all $i \in \mathbb{N}_n$.

Within probability theory full certainty is expressed by $p(s) = 1$ for a particular $s \in S$. The expression of total ignorance in probability theory, on the other hand, is given by

$$p(s) = \frac{1}{|S|}$$

for all $s \in S$.

The set of outcomes of throwing a fair die can be represented by the universal set $X = \{x_1 =$ ⚀$, x_2 =$ ⚁$, x_3 =$ ⚂$, x_4 =$ ⚃$, x_5 =$ ⚄$, x_6 =$ ⚅$\}$ where we see x_i has a value of i that corresponds to the numbers of spots on the upper side of the die when it is thrown.

If the die is absolutely fair then all of the six outcomes are equally likely and $p_i = 1/6$ for $i \in \mathbb{N}_6$ since the six probabilities must sum to one. We can now talk about the probability of any outcome of rolling the die. The probability that a thrown die is even is $P(E)$ where $E = \{x_2, x_4, x_6\}$ and

$$\begin{aligned} P(E) &= P(\{x_2, x_4, x_6\}) \\ &= p_2 + p_4 + p_6 \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{2} \end{aligned}$$

so that the probability of an even throw is one-half. The probability of throwing a number greater than four is $P(G)$ where $G = \{x_5, x_6\}$ and

$$\begin{aligned} P\left(G\right) &= P\left(\{x_5, x_6\}\right) \\ &= p_5 + p_6 \\ &= \tfrac{1}{6} + \tfrac{1}{6} \\ &= \tfrac{1}{3} \end{aligned}$$

so that probability of throwing a number greater than four is one-third.

Another important formula of probability is due to the fact that A and not A include everything, and have nothing in common, thus $A \cup \bar{A} = S$ and $A \cap \bar{A} = \emptyset$. Since the probability of everything (the sample space) is one (See Eq. 4.2) and the probaility of the empty set is zero (See Eq. 4.1) we can use Eq. 4.3 to conclude that the probability of A union not A is one.

$$\begin{aligned} P(A \cup \bar{A}) &= P(A) + P(\bar{A}) - P(A \cap \bar{A}) \\ P(S) &= P(A) + P(\bar{A}) - P(\emptyset) \\ 1 &= P(A) + P(\bar{A}) - 0 \\ 1 &= P(A) + P(\bar{A}) \end{aligned}$$

If we rearrange $P(A) + P(\bar{A}) = 1$ we get the following rule:

$$P(\bar{A}) = 1 - P(A)\,.$$

A similar derivation using the facts that $(A \cap B) \cup (A \cap \bar{B}) = A$ and $(A \cap B) \cap (A \cap \bar{B}) = \emptyset$ gives a useful formula:

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \tag{4.4}$$

4.1.1. Conditional Probability

The notation for conditional probabilites is $P(A \mid B)$ and the formula for their computation is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}\,. \tag{4.5}$$

Let us consider rolling a dice twice. There are 36 possible outcomes since there are six differnt outcomes for the first die and six differnt outcomes for the second die. If the die is fair each individual throw has six outcomes of equal probability, and that probability is $\frac{1}{6}$. The two throws are independant so that each of the 36 compound events, such as tossing a ⚃on the first throw and a ⚁on the second has probability $\frac{1}{36}$. This follows from the same logic as the head and tails Example above. Let A be the event that the sum of the spots on the two rolls is seven. Let B be the event that one of the two rolls is a five. Specifically

$A = \{$⚀⚅, ⚁⚄, ⚂⚃, ⚃⚂, ⚄⚁, ⚅⚀$\}$

and

$B = \{$⚄⚀, ⚀⚄, ⚄⚁, ⚁⚄, ⚄⚂, ⚂⚄, ⚄⚃, ⚃⚄, ⚄⚄, ⚄⚅, ⚅⚄$\}$

. To apply formula (4.5) we will aslo need to know the probability of $A \cap B = \{$⚁⚄, ⚄⚁$\}$. Now the probability of $A \cap B$ is $\frac{1}{36} + \frac{1}{36} = \frac{2}{36}$ and the probability of B is $\frac{11}{36}$ (which is $\frac{1}{36}$ added 11 times). We can now calculate the probability one of the dice was a five given that the sum of the two rolls was seven, or $P(A \mid B)$.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{36}}{\frac{11}{36}} = \frac{2}{11}\,.$$

Example 21. Suppose the sample space is the set of adults in a small town who have completed the requirements for a college degree. We shall classify these adults by employment and sex. The resultant data is summarized in the folwing Table.

	Employed	Unemployed	Total
Female	140	260	400
Male	460	40	500
Total	600	300	900

We will be interested in two events:

M: a man is chosen

E: an employed adult is chosen

Let us calculate the probability that someone is male given that they are employed. There are 900 adults under consideration and 460 are Male and Employed so $P(M \cap E) = \frac{460}{900}$. There are 600 employed adults so $P(E) = \frac{600}{900}$. Finally we calculate that

$$P(M \mid E) = \frac{P(M \cap E)}{P(E)} = \frac{\frac{460}{900}}{\frac{600}{900}} = \frac{23}{30}\,.$$

4.2. Independence

Example 22. If we rearrange formula (4.5) we get the multiplication rule for probabilites:

$$P(A \cap B) = P(A)P(B \mid A) \tag{4.6}$$

Two events A and B are independent if $P(A)$ does not depend on the occurance of B and $Pr(B)$ does not depend on the occurance of A. In other words, if $P(A) = P(A \mid B)$ and $P(B) = P(B \mid A)$ then A and B are independant, otherwise they are dependant. If events A and B are independent, then the probability of A and B both occuring, $P(A \cap B)$, is just the product of the individual probabilities of A and B:

$$P(A \cap B) = P(A)P(B)\,.$$

4.3. Joint Distributions

When a probability distribution function p is defined on a Cartesian product $X \times Y$, it is called a *joint distribution*. In this case we have that $p : X \times Y \to [0,1]$. The associated marginal distributions are determined by the formulas

$$p_X(x) = \sum_{y \in Y} p(x, y) \tag{4.7}$$

for each $x \in X$ and

$$p_Y(y) = \sum_{x \in X} p(x, y) \tag{4.8}$$

for each $y \in Y$. The *noninteraction* of the marginal bodies of evidence is defined by the condition

$$p(x, y) = p_X(x) \cdot p_Y(y) \tag{4.9}$$

for all $x \in X$ and all $y \in Y$.

Example 23. Suppose we flip a fair coin twice. Let H denote "heads" and T denote "tails." The sample space consists of the four possible outcomes: $S = \{HH, TH, HT, TT\}$. We assume the coin is fair so that each of the four outcomes has an equal probability. We will now compute these probabilities. For a fair coin the probability of "heads" on any flip equals the probability of "tails" and since these values must add up to one we have that:

$$p(H) = p(T) = 1/2.$$

The first flip does not depend on the occurance of the second flip and the second flip does not depend on the first flip. This is what we mean by the independence of the events "first flip" and "second flip". We denote "first flip" as smple one or $s1$ and "second flip" as sample 2 or $s2$. The probability of any $s \in S$ is then

$$p(s1s2) = p(s1)p(s2)$$

for $s1, s2 \in H, T$. Specifiicall, when $s1 = H$ and $s2 = H$, then

$$p(HH) = p(H)p(H) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Clearly then, $p(HH) = p(TH) = p(HT) = p(TT) = \frac{1}{4}$. The four events HH, TH, HT, and TT are exhaustive (ther are no other possible outcomes) and mutually exclusive, no two can occur at the same time. As a check on the rule of probability given by Equation ((4.2)) we note that

$$P(S) = p(HH) + p(TH) + p(HT) + p(TT) = 1.$$

Example 24. We can define other events that can occur as the result of flipping a coin twice. For example, the event that the coin is gives the same reult on both flips is $A = \{HH, TT\}$. The event that the second flip is tails is $B = \{HT, TT\}$. The event that both flips are tails is $C = \{TT\}$ .We can then calculate the probabilities of these events:

$$P(A) = P(\{HH, TT\}) = p\,(HH) + p\,(TT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(B) = P(\{HT, TT\}) = p\,(HT) + p\,(TT) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(C) = P(\{TT\}) = p\,(TT) = \frac{1}{4}\,.$$

Example 25. 20% of College students are seniors. 61% of College students are female. If we assume theer is no interaciton then the probability that a randomly chosen student is a senior female is %12.2.

4.4. Bayes Theorem

Bayes Theroem allows us to calculate the difference between prior (before the fact) and posterior (after the fact) knowledge. The formula is easy to derive by noting that $P(A \cap B) = P(A)P(B \mid A)$ and that $P(A \cap B) = P(B)P(A \mid B)$. If we apply a little algebra to the equivalence $P(A)P(B \mid A) = P(B)P(A \mid B)$ we get Bayes Theroem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{4.10}$$

Example 26. Suppose we continue using the data of Example 21 and ask "What is the probability that someone is employed given that they are male?" We have already calculated the probability that someone is male given that they are employed, or $P(M \mid E)$. There are 500 adult males so $P(M) = \frac{500}{900}$. There are 600 employed adults so $P(E) = \frac{600}{900}$. Bayes Theroem now gives:

$$\begin{aligned} P(E \mid M) &= \frac{P(M \mid E)P(E)}{P(M)} \\ &= \frac{\frac{23}{30} \times \frac{600}{900}}{\frac{500}{900}} \\ &= \frac{23}{25} \end{aligned}$$

which makes sense since we already knew that 460 out of 500 males were emplyed.

Where Bayes Thereom becomes really important is when probability is viewed as rational coherent degrees of belief. The Bayes Therorem allows us to update probabilities as new information becomes available. Since probabilities are beleifs, we can start with a subjective estimate. Thus we can start with a beleif that the probability of the Detroit Lions winning Superbowl XLIII is $\frac{1}{100}$. As they start to win games, we can look at the records of previous seasons to calculate how often a team with an identical record went on to win the superbowl. This allows us to update the odds.

Example 27 (Drug testing)**.** Suppose we run a company and wish to test new employees for drug use. We have a test that that is fairly accurate in that it is positive for drug users 95% of the time and negative for non-drug usere 95% of the time. Suppose that 1% of the employees are drug users. Let us calculate the probability that the person is not a drug user given that the test is positive. Let T be the event that the test is positive. Let D be the event that the employee is a drug user. Here is what we know:

$P(D)$ 1% of the employees use drugs

$P(\bar{D})$ 99% of the employees do not use drugs

$P(T : \bar{D})$ 5% of the non–drug users test positive

$P(\bar{T} : \bar{D})$ 95% of the non–drug users test negative

$P(T : D)$ 95% of the drug users test positive

$P(\bar{T} : D)$ 5% of the drug users test negative

We can use Eq. 4.4to calcualte the probability positive test result $P(T) = P(T \cap D) + P(T \cap \bar{D})$. By Eq. 4.6$P(T \cap D) = P(T \mid D)P(D)$ or $P(T \cap D) = 0.95 \times 0.01$. Similarly $P(T \cap \bar{D}) = P(T \mid \bar{D})P(\bar{D})$ or $P(T \cap \bar{D}) = 0.05 \times 0.99$.

By Bayes Theroem

$$\begin{aligned} P(\bar{D} \mid T) &= \frac{P(T \mid \bar{D})P(\bar{D})}{P(T)} \\ &= \frac{0.05 \times 0.99}{0.95 \times 0.01 + 0.05 \times 0.99} \\ &= 0.838983051 \end{aligned}$$

or about 84%.

Sometimes Bayes Theorem is presented in terms of an hypothesis *h* and evidence *e*,

$$P_e(h) = \frac{P_e(h)P(e)}{P(h)}\,. \tag{4.11}$$

This formula tells how to recalculate the probability of our hypothesis as new evidence is accumulated.

4.5. Continuous Sample Spaces

For a continuous universal set S the definitions of probability distribution and probability measures are more complex. It is usually necessary to limit the the set of events to those that are well behaved mathematically. The probability measure is usually defined on a Borel set and the formula for the union of two sets needs to be expanded to one suitable for countable unions of sets. This book will have little need of these precise definitions. What is important about continuous probability distributions is that instead of summing the distributional values as is done for discrete distributions we must integrate. A probability distribution is still a function from the sample space into the unit interval, but instead of summing to one we have . The formula for the probability of an event A is now given by

$$P(A) = \int_{x \in A} p(x)\ dx$$

where A is a subset of the continuum S.

4.6. Statistics

A random variable is a real valued function whose domain is a sample space.

A statistic is a number that describes some property of a random variable. As an example let X be a random variable that maps the set of outcomes of rolling a die onto the numbers one through six. Thus the sample space is $S = \{s_1, s_2, s_3, s_4, s_5, s_6\}$ and $X : S \to \{1, 2, 3, 4, 5, 6\}$ is specified by $X(s_i) = i$ for $i \in \mathbb{N}_6$. Let $x_i = X(s_i)$. If we repeatedly roll the die, and keep track of the outcome, we can calculate a statistic, such as the average value of the outcome, that gives us information about the whole experiment, and hence about the die itself. If we roll the die thirty times then we might get five ones, six twos, three threes, six fours, six fives and four sixes. We can calculate the average by adding up all the outcomes and dividing by thirty, giving 3.4667 for the average.

The average is a measure of central tendency. In statistics the three most common measures of central tendency are the mean, the median and the mode. The mean is just the average value. The median is the middle number, the number that half the data is above and half the data is below. The mode is the outcome that has the greatest frequency.

Besides the measures of central tendency, measures of dispersion are the most widely used statistics. The variance is the average value of the square of the distance of the random variable from the mean, for the die example given above the variance is 2.8489. The standard deviation is the square root of the variance.

Of course there is always some difference between the real and the ideal. One would be shocked if the all thirty rolls of the die produced a one! In fact there is always some difference between an experimental mean, usually symbolized by m, and an ideal mean, usually represented by μ.

The expected value of a finite random variable, $E(X)$, is the sum of the probability of that outcome multiplied by the corresponding value of the random variable,

$$\begin{aligned} E(X) &= \sum_{s \in S} p(s) \cdot X(s) \\ &= \sum_{i=1}^{n} p_i \cdot x_i \end{aligned}$$

where, n, is the size of the universe and for any function $f : X \to \mathbb{R}$.

$$\begin{aligned} E(f(X)) &= \sum_{s\in S} p(s) \cdot f(X(s)) \\ &= \sum_{i=1}^{n} p_i \cdot f(x_i) \end{aligned}$$

In the case of a fair die the ideal mean is the expected value of the die

$$\begin{aligned} \mu &= E(X) \\ &= (1 \cdot \frac{1}{6}) + (2 \cdot \frac{1}{6}) + (3 \cdot \frac{1}{6}) + (4 \cdot \frac{1}{6}) + (5 \cdot \frac{1}{6}) + (6 \cdot \frac{1}{6}) \\ &= 3.5 \end{aligned}$$

The ideal variance, σ^2, is the expected value of the square of the distance of x_i from the expected mean.

$$\begin{aligned} \sigma^2 &= E((X-\mu)^2) \\ &= ((1-3.5)^2 \cdot \frac{1}{6}) + ((2-3.5)^2 \cdot \frac{1}{6}) + ((3-3.5)^2 \cdot \frac{1}{6}) + \\ &((4-3.5)^2 \cdot \frac{1}{6}) + ((5-3.5)^2 \cdot \frac{1}{6}) + ((6-3.5)^2 \cdot \frac{1}{6}) \\ &= 2.917 \end{aligned}$$

The standard deviation is the square root of the variance.

The expected value of the surprise, $-\log_2 p_i = \log_2 \frac{1}{p_i}$ is called the Shannon entropyShannon (1948) and is the basis of information theory in statistical applications:

$$S(p) = -\sum_{i=1}^{n} p_i \log_2 p_i \ .$$

For the continuous case the formulas for expected values become ,

$$E(X) = \int_S X(s) \cdot p(s)\, ds$$

where, n, is the size of the universe and for any function $f : X \to \mathbb{R}$.

$$E(f(X)) = \int_S f(X(s)) \cdot p(s)\, ds$$

4.7. Homework

Let us define the universal sets

$$X = \{1, 2, 3, 4, 5, 6\}, \tag{4.12}$$

$$Y = \{a, b, c\}, \tag{4.13}$$

Le

$$\begin{aligned}
A &= \{1, 2\}, \\
B &= \{1, 3\}, \\
C &= \{2, 4, 6\}, \\
D &= \{a, b\}, \text{ and} \\
S &= \{\langle 1, a\rangle, \langle 2, b\rangle, \langle 3, b\rangle, \langle 2, c\rangle\}.
\end{aligned}$$

Given the above information, answer the following questions.

1. If we have the uniform probability distribution on X so that $p(x) = \frac{1}{6}$ for all elements x of X, what is $P(A)$, $P(B)$, $P(C)$, $P(\emptyset)$, and $P(X)$?
2. What is the expected value of x under the assumptions of the previous question?
3. If we have a probability distribution p on X so that $p(1) = 0.1$, $p(2) = 0.2$, $p(3) = 0.3$, $p(4) = 0.1$, $p(5) = 0.2$, and $p(6) = 0.1$ then what is $P(A)$, $P(B)$, $P(C)$, $P(\emptyset)$, and $P(X)$?
4. What is the expected value of x under the assumptions of the previous question?
5. What is the probability of getting lung cancer if a person is a smoker.

Part II.

Fuzzy Sets

5. Fuzzy sets

5.1. Introduction

Vaugeness is a pervasive part of the human experience. Human language is an imprecise tool. Human perception is fraught with inaccuracy. Memories are fleeting and malleable. The real world is not an abstraction; it is not clearly perceived, well defined, and precisely calculated.

The advent of computer technology has made philosophical dilemmas moot. Programmers are tasked with the creation of software that works, for and with humans. Trying to bridge the cybernetic gap has led to the creation and use of a host of technologies — artificial intelligence, data mining, expert systems, for example — that need to represent and manipulate the uncertainties of real life.

One of the most powerful tools of the cybernetic age is fuzzy set theory. A fuzzy set has graded membership. Thus it is designed to handle vagueness.

For many years, Pluto was a planet and then in 2006, it was not. Pluto had not changed. What had changed was its classification. The discovery of many distant objects in the solar system, some larger than Pluto, had caused astronomers to question the classification of Pluto. In fact, on August 24, 2006, the International Astronomical Union defined the term “planet” for the first time! Pluto did not fit the definition. Pluto is now classified as a “minor planet”.

Fuzzy set theory takes a different approach. The classifications *planet* and *minor planet* would not have distinct boundaries. Instead, some objects might have characteristics that allow for partial inclusion in both categories. It would place Pluto in the set of planets but not to the same degree as the eight inner planets.

It is also very important to notice that up until 2006 it would seem that the scientific term planet was vague! Was it that scientists did not know what they were talking about? That is not the case. The lack of a precise definition of planet in no way impacted its use or usefulness. Historically, the planets were Mercury, Venus, Earth, Mars, Jupiter and Saturn, (and, according to Ptolemy and others, the Moon). Galileo’s discovery of Jupiter’s moons eliminated the Moon from the list and the telescope eventually added Uranus and Neptune. Finally, Pluto was discovered and added even though it turned out to be smaller than the Moon. The discovery of objects orbiting other suns brought the lack of definition of planet into sharp focus. Something had to be done.

For computers to process vague and ill defined information, something had to be done.

5.2. Set Theory

As explained in Chapter 2, sets are the basis of all mathematics. A set is a collection of objects. The set of candidate objects is the universal set, most often labeled X or U. Typical universal sets are the real numbers or the natural numbers.

The important aspect in defining a set is that the definition must enable one to determine which objects are in the set and which objects are not in the set. Sets are specified using one of three methods: explicit listing of elements, specification of a necessary property, and the use of characteristic functions.

1. For finite sets one can use an complete listing such as $A = \{2, 4, 6\}$. For countably infinite sets the ellipsis "$\ldots$" denotes "and so on" as in the set of all natural numbers $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$.

2. Specifying a set by listing a property that all its contents must have can be used for any type of set. For example we can have $A = \{x \mid x$ is an even natural number less than $7\}$ which is the same A as in the previous item. Another example is the infinite set $P = \{x \mid x$ is a prime number$\}$.

3. For the purposes of this book the characteristic function method is the most important way of determining a set's contents. The characteristic function of a set A, $\chi_A(x)$, is zero if x is not in the set A and one if x is in the set A.

$$\chi_A(x) = \begin{cases} 1 & x = 2, 4, 6 \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

5.3. Fuzzy Sets

Consider the set F of delicious foods. Traditional set theory says that every candidate food x must be either in the set or not in the set. Most people would place chocolate in the set F but how about caviar. Caviar is salty and fishy and not to everyone's taste, yet those who do like it consider it a delicious food. There is a famous adage that "the best sauce is hunger," and to a hungry man a simple slice of bread and bowl of soup will seem very delicious indeed. The set F suffers because delicious food is not an easily defined notion, as opposed to the notion of a prime number which can be defined precisely.

Computers can understand things that can be defined precisely, and have a lot of trouble with concepts that humans grasp in childhood. Take the set B of all balls. Certainly ping-pong balls and baseballs are in B, and most people would admit that a football, while not precisely round, is a ball. But how about a whiffleball or the shuttlecock used in badminton?

The characteristic functions of classical set theory maps elements of some universal set X into the binary set $\mathbb{B} = \{0, 1\}$. This dichotomy is typical of classical western thinking, whether it is as *yes* and *no*, *true* and *false*, or *one* and *zero*. The fundamental idea of fuzzy set theory is that real world phenomena cannot be divided easily into such *black* and *white* divisions. For example, what is, exactly, the dividing line between *rich* and *poor*. Where does the middle class fit into such a categorization? Is every day either *sunny* or *cloudy*? Is all food either *good* or *bad*?

In fuzzy set theory we extend the image set of the characteristic function from the binary set $\mathbb{B} = \{0, 1\}$ which contains only two alternatives, to the unit interval $\mathbb{U} = [0, 1]$ which has an infinite number of alternatives. We even give the characteristic function a new name, the membership function, and a new symbol μ, instead of χ. This introduces a richer and highly applicable field which measures the world in shades of gray. It measures wealth in gradations that include "upper middle class," it represents weather with values that include "patchy clouds," and even allows us to classify some as being "just okay."

The mathematics of fuzzy set theory is often more difficult than that of the traditional set theory because the continuous interval $\mathbb{U} = [0, 1]$ is inherently more complex than the binary set $\mathbb{B} = \{0, 1\}$.

5.4. Membership functions

The terminology *fuzzy* is due to Lotfi Zadeh who created the field of fuzzy sets in Zadeh (1965). Zadeh knew that there were other kinds of uncertainty in the world besides the randomness that is handled in probability theory. Things that are not random can still be uncertain. If someone places a die on the floor on the opposite side of the room then there is no randomness involved in ones inability to precisely determine the number of spots on the top of the die. Zadeh's motivation came primarily from modeling human language. The statement "Sally is very tall" contains ambiguities and imprecisions that have nothing to do with randomness.

Zadeh decided to term this kind of uncertainty "fuzzy" and we have inherited his terminology.

Since its inception there has been great criticism of fuzzy set theory leading to fuzzy thinking. One might as well criticize probability theory for leading to random actions. Probability theory is a precise mathematical device for processing data whose source is a random event. Fuzzy sets are a precise mathematical tool for processing data that is derived from vague sources. Human beings transmit vague information such as "Juan is just a teenager."

In fact, if you ask a question like, "How old is Juan?" to most people, you will not get a numeric answer. Only his friends and relatives usually know his exact calendar age. An acquaintance will answer that he is a "teenager" or "young" or "adolescent." And the age of the person answering the question will greatly influence the answer given. A person of seventy might say Juan is a "boy" whereas a contemporary would not. A contemporary might even say "He is the same age I am," which provides no direct numerical information.

The vagueness of language, and its mathematical representation and processing, is one of the major areas of study in fuzzy set theory.

Since there is nothing fuzzy about a fuzzy set we must be specific about its definition and interpretation. A fuzzy set is just a function. Its domain is some universal set X. Its range is the unit interval $\mathbb{U} = [0, 1]$.

When Zadeh (Zadeh (1965)) originated fuzzy sets he introduced the *membership function* of a fuzzy set A and used the notation

$$\mu_{\mathsf{A}} : X \to [0, 1] \, .$$

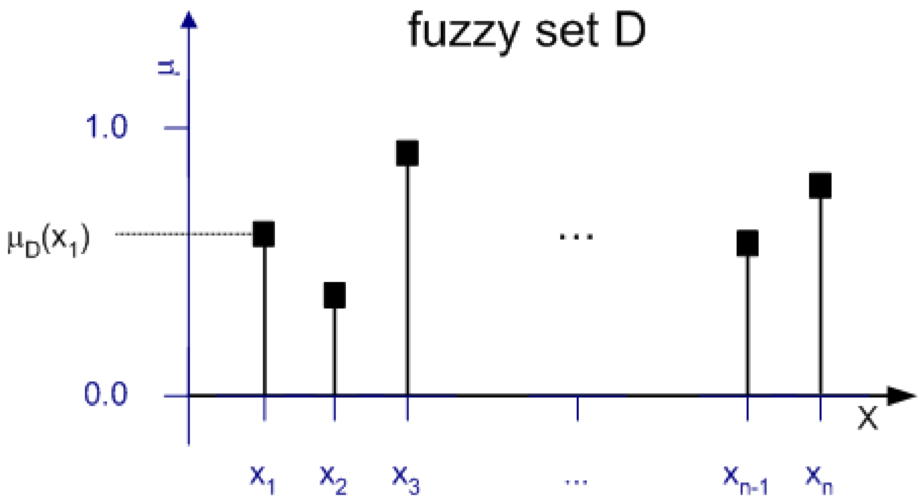


Figure 5.1.: A fuzzy set D on a discrete universe $X = \{x_1, x_2, ..., x_n\}$.

Specifying the membership function specifies the fuzzy set. This notation and definition is extremely similar to that of the characteristic function of traditional or crisp set theory. Characteristic functions use the Greek letter chi, χ, instead of mu, μ, and limit the function to zero and one while membership functions allow the complete spectrum from zero to one.

Example 28. Let $X = \{a, b, c\}$ and define the fuzzy set A as follows, set $\mu_{\mathsf{A}}(a) = 1.0$, $\mu_{\mathsf{A}}(b) = 0.7$, and $\mu_{\mathsf{A}}(c) = 0.4$. Thus a is completely compatible with the label or classification A while b is only very compatible. On the other hand c is somewhat incompatible with the notion conceptualized by A.

Given this definition of μ_{A}, A is now a fuzzy set and we possess an explicit definition of its membership function. Unfortunately the tag of the fuzzy set, A, is difficult to read as a subscript, and the membership function of a fuzzy set with a subscript, such as A_i with membership function μ_{A_i}, compounds the subscript problem.

As observed in Chapter 2 it is often desirable to use the same tag for a set and for its characteristic function. We will expand on that observation here in this section, by using the tag A to represent both the fuzzy set and its membership function. Thus, if A is a fuzzy set then we will also use A as the label of a function from a universe of discourse X into the unit interval $\mathbb{U} = [0, 1]$

$$\mathsf{A} : X \to [0, 1] . \tag{5.2}$$

Example 29. The membership function of the fuzzy set in the previous example (#28) is specified as $\mathsf{A}(a) = 1.0$, $\mathsf{A}(b) = 0.7$, and $\mathsf{A}(c) = 0.4$. In addition we might use the notion that a function maps a element x to a value $f(x)$ to present A as a set of ordered pairs of elements of X and their associated membership grades, $\langle x, \mathsf{A}(x)\rangle$, so that $\mathsf{A} = \{\langle a, 1.0\rangle, \langle b, 0.7\rangle, \langle c, 0.4\rangle\}$. Finally, we might present the values as a table:

$$\begin{array}{c|c} x & \mathsf{A}(x) \\ \hline a & 1.0 \\ b & 0.7 \\ c & 0.4 \end{array} \tag{5.3}$$

Example 30. Let $X = \{al, bo, cam\}$ and define the fuzzy set Tall as follows, set $\mu_{\mathsf{Tall}}(al) = 1.0$, $\mu_{\mathsf{Tall}}(bo) = 0.7$, and $\mu_{\mathsf{Tall}}(cam) = 0.4$. Thus al is completely Tall while bo is somewhat Tall. On the other hand cam is not very Tall.

It is probably useful to point out that unlike probability, the notation used in books and papers about fuzzy set theory is not standardized. The notation for fuzzy set theory is different in almost every paper that one tries to read. (Fuzzy set theory is a young discipline.) A fuzzy set was traditionally indicated by the tilde above the label, such as $\tilde{A}$, that indicated that A was indeed a fuzzy set. Thus much of the early, and current, literature includes phrases such as "the fuzzy set $\tilde{E}$". It is more common, however, for he current literature to omit the tilde. Besides the notation $\mathsf{A}(x)$ that this book will use, common notation for a fuzzy set membership function include: $\mu_{\mathsf{A}}(x)$, $m_{\mathsf{A}}(x)$, $\chi_{\mathsf{A}}(x)$, $\mu_{\tilde{A}}(x)$, $m_{\tilde{A}}(x)$, $\chi_{\tilde{A}}(x)$, and $\tilde{A}(x)$. Common synonyms for membership function are membership grade, compatibility index, and characteristic function.
[1]

The statement $\mathsf{A}(b) = 0.7$ is interpreted as saying that the membership grade of b in the fuzzy set A is seven-tenths. The only difference between a traditional set and a fuzzy set is the image of their membership functions. A traditional set has its membership grades taking values in the set $\{0, 1\}$ while a fuzzy set has its membership grades in the unit interval $[0, 1]$. A standard set will be called a crisp set whenever it is necessary to distinguish it from a fuzzy set. The universal set X is always a crisp set.

The difference between the image sets $\mathbb{U} = [0, 1]$ and $\mathbb{B} = \{0, 1\}$ means that fuzzy sets do not view the world in black and white but instead see the world in shades of gray. All humans (except ideologues) know that the world is not made up of absolutes but traditional mathematics is composed of idealized abstractions. This has forced divisions that are not natural, and more important, these divisions are contrary to the way humans represent and process information. For example consider a crisp set B and a fuzzy set B both of which are to represent the concept blue. An object b, such as a shirt in your closet, is either in B or it is not. Suppose that b_1 is a tartan plaid shirt, b_2 is a white shirt, and b_3 is a blue and green pinstripe shirt. For the crisp

[1] One particular notation that is outdated is Zadeh's original fraction notation. It has the advantage of being horizontal, which takes up less space, and is was easy to type back in the days before word processors. This notation is a list of fractions that use + as a separator, but the + does not represent addition, it simple separates terms. Each fraction has a denominator which is the element of the fuzzy set under discussion and a numerator that specifies the membership value. Thus the fuzzy set A specified by $\mathsf{A}(a) = 1.0$, $\mathsf{A}(b) = 0.7$, and $\mathsf{A}(c) = 0.4$ is given as

$$\mathsf{A} = \frac{1.0}{a} + \frac{0.7}{b} + \frac{0.4}{c}\,. \tag{5.4}$$

We reiterate here that in the presentation of A in Eq. 5.4 the fraction does not indicate division and the plus sign does not indicate addition! Instead the element below the bar has the membership grade equal to the value above the bar and the plus sign only separates members of the list.

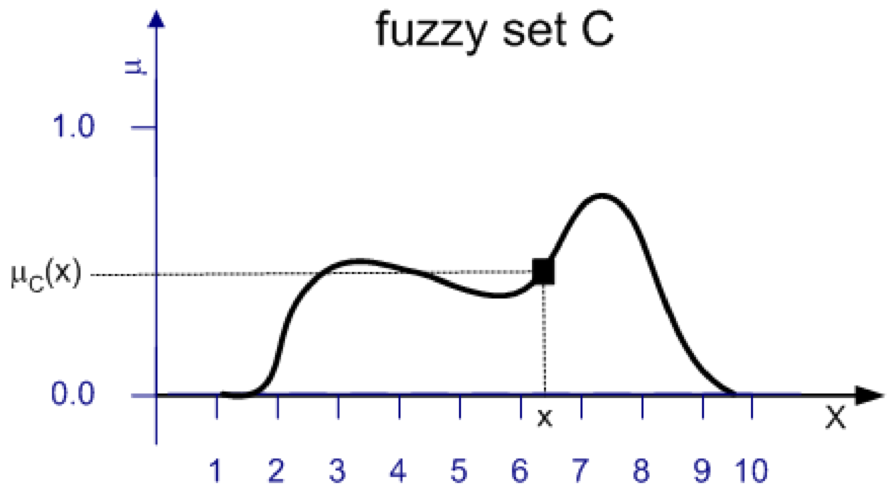


Figure 5.2.: A fuzzy set C on a continuous domain $X = [1, 10]$.

set B we must make a decision on which of these objects are ideally blue and which are not. On the other hand the fuzzy set B allows for a graded membership that can allow b_1, b_2, and b_3 partial membership based on how much blue they contain, or how much blue that a specific observer says that the shirts contain. The fuzzy set is much more in line with the human representation. If a shirt is needed to match a pair of blue-jeans then it may be that b_3 is the best choice even if it would not have been accepted for membership in the crisp set B of blue objects.

5.5. Fuzzy Set Operations

In this section we will define what it means for a fuzzy set to be contained in another fuzzy set. We will also provide operators for fuzzy sets that correspond to "and" "or", and "not" of human logic. We will define a fuzzy intersection to represent "and", a fuzzy union to represent "or", and a fuzzy complement to represent "not". In fact we will use the exact notation for fuzzy intersections, unions, and complements that we used for crisp intersections, unions, and complements. This usage can be justified by noticing that crisp sets are, in a sense, a special case of fuzzy sets. Both, crisp sets and fuzzy sets can be defined using membership functions. The crisp set membership values of zero and one are contained in the unit interval and crisp sets can be thought of as fuzzy sets with a restricted image set. The definition of a fuzzy intersection will be crafted so that if A and B have $\{0, 1\}$ as their image set, and thus have no fuzziness to them, then the fuzzy intersection of A and B behaves precisely like the intersection of crisp sets. This will also be true for the fuzzy union, fuzzy complement, and the fuzzy subset relation. Because of this we do not need to keep using the prefix "*fuzzy*" to these operations. The fuzzy intersection is the crisp intersection whenever the sets involved correspond to crisp sets, that is, have no fuzziness, no elements with partial grades of membership.

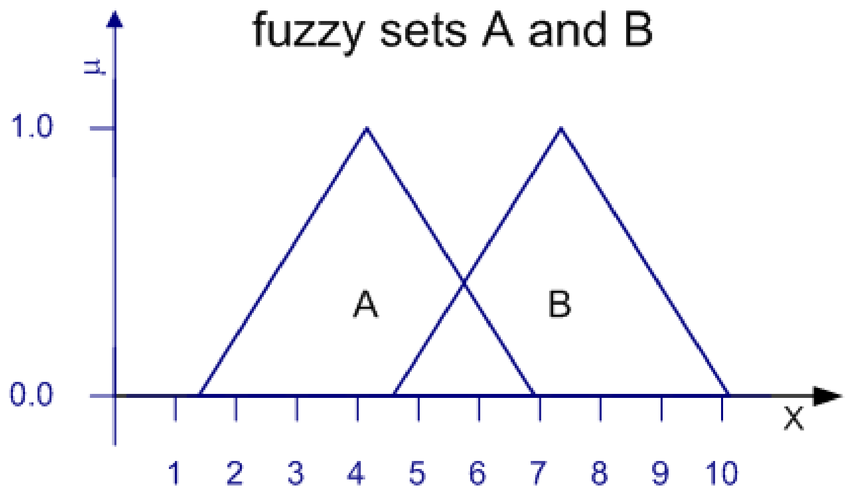


Figure 5.3.: Fuzzy sets A and B.

All the definitions for complements, unions, and intersections are given in terms of membership functions. Thus to define a fuzzy complement A^c we define its membership function in terms of the membership function of the fuzzy set A.

5.5.1. Subsets

A fuzzy set A is a subset of a fuzzy set B if $\mathsf{A}(x)$ is less than or equal to $\mathsf{B}(x)$ for all x in X. In crisp set theory a subset has less things, while in fuzzy set theory a subset has things to a lesser degree.

Definition 6 (Subset)**.** Thus $\mathsf{A} \subseteq B$ if and only if

$$\forall x \in X \ \mathsf{A}(x) \leq \mathsf{B}(x). \tag{5.5}$$

For A to be equal to B the membership values of $\mathsf{A}(x)$ and $B(x)$ must be equal for all x.

Example 31. If we have the fuzzy set A,

$$\begin{array}{c|c} x & \mathsf{A}(x) \\ \hline a & 1.0 \\ b & 0.7 \\ c & 0.4 \end{array} \tag{5.6}$$

from Example 29 then $\mathsf{D} = \{\langle a, 0.9\rangle, \langle b, 0.0\rangle, \langle c, 0.4\rangle\}$,

$$\begin{array}{c|c} x & \mathsf{D}(x) \\ \hline a & 0.9 \\ b & 0.0 \\ c & 0.4 \end{array} \tag{5.7}$$

is a fuzzy subset of A since:

$$\begin{array}{c} \mathsf{A}(a) = 1.0 \geq 0.9 = \mathsf{D}(a) \\ \mathsf{A}(b) = 0.7 \geq 0.0 = \mathsf{D}(b) \\ \mathsf{A}(c) = 0.4 \geq 0.4 = \mathsf{D}(c) \end{array}$$

This is easier to see if we use a table to show both fuzzy sets:

$$\begin{array}{c|ccc} x & \mathsf{D}(x) & \mathsf{A}(x) & \mathsf{D}(x) \leq \mathsf{A}(x) \\ \hline a & 0.9 & 1.0 & 1 \\ b & 0.0 & 0.7 & 1 \\ c & 0.4 & 0.4 & 1 \end{array} \tag{5.8}$$

then $\mathsf{D}(x) \subseteq \mathsf{A}(x)$ if and only if the final comparison column is all ones, in computer science $1 \equiv True$ and $0 \equiv False$.

Definition (Set equality)**.** $A = B$ if and only if

$$\forall x \in X \ \mathsf{A}(x) = \mathsf{B}(x). \tag{5.9}$$

It is simple to show that $\mathsf{A} \subseteq \mathsf{B}$ and $\mathsf{A} \supseteq \mathsf{B}$ imply that $\mathsf{A} = \mathsf{B}$.

The class of all fuzzy subsets of the universe X is called the fuzzy power set. There are an infinite number of fuzzy subsets of any non–empty universe.

The class of all fuzzy sets defined upon the universal set X is called the fuzzy power set.

Definition 7 (Fuzzy power set)**.** Suppose that X is a crisp universal set, let the class of all fuzzy sets defined upon X be denoted $\mathcal{F}(X)$ and called the fuzzy power set.

Thus $\mathcal{P}(X)$ is the crisp power set and $\mathcal{F}(X)$ is the fuzzy power set.

Unfortunately it is impossible to give a simple example of a fuzzy power set since it is infinite in size. Since we can not list all the real numbers between zero and one, we can not list all the fuzzy subsets of a crisp set $X = \{a, b, c\}$.

5.5.2. Intersection

The intersection of two fuzzy sets was defined by Zadeh to be a fuzzy set $\mathsf{C}(x) = (\mathsf{A} \cap \mathsf{B})(x)$ with membership function $\mathsf{C}(x)$ equal to the minimum of the values of the membership grades of x in A and B.

Definition 8 (Fuzzy intersection)**.** The intersection of fuzzy sets A and B is the fuzzy set $\mathsf{A} \cap \mathsf{B}$ with a membership grade for every $x \in X$ given by

$$(\mathsf{A} \cap \mathsf{B})(x) = \min[\mathsf{A}(x), \mathsf{B}(x)] \tag{5.10}$$

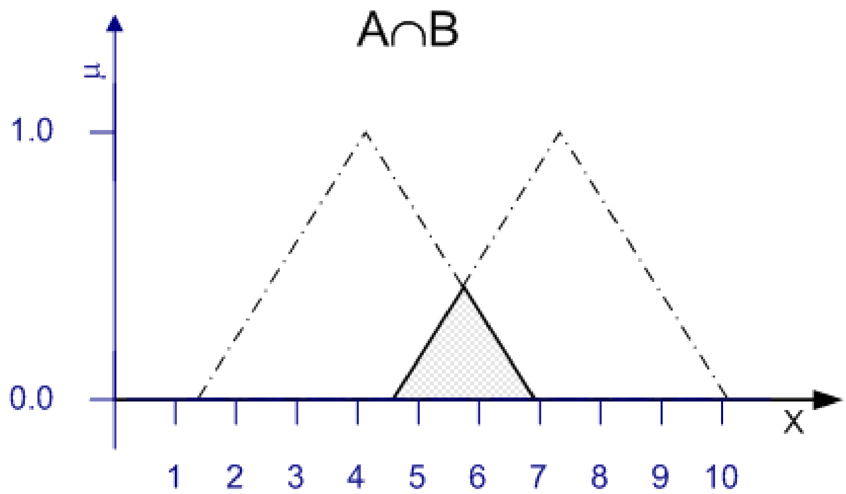


Figure 5.4.: $A \cap B$, the intersection of fuzzy sets A and B.

for all x in X. The minimum operator is represented by the symbol "$\wedge$"so that $\min[a, b]$ can also be written $a \wedge b$ and the intersection membership function is often written as

$$(\mathsf{A} \cap \mathsf{B})(x) = \mathsf{A}(x) \wedge \mathsf{B}(x)\,. \tag{5.11}$$

Example 32. Consider the fuzzy set A and B given in the following table:

x	A	B
a	1.0	0.3
b	0.7	0.6
c	0.4	0.9

(5.12)

The intersection of two fuzzy sets A and B is a fuzzy set that contains an element x of X at the minimum membership degree consistent with the membership grade of x in A and of x in B. Since c is in A to the degree 0.4 and c is in B to the degree 0.9 we know that c is in the intersection of the fuzzy sets A and B to the degree 0.4 since this is the minimum of the two membership values. The complete result is the fuzzy set A ∩ B with membership values as follows:

x	A ∩ B
a	0.3
b	0.6
c	0.4

(5.13)

5.5.3. Union

The union of two fuzzy sets was defined by Zadeh to be a fuzzy set $\mathsf{C}(x) = (\mathsf{A} \cup \mathsf{B})(x)$ with membership function $\mathsf{C}(x)$ equal to the maximum of the values of the membership grades of x in A and B.

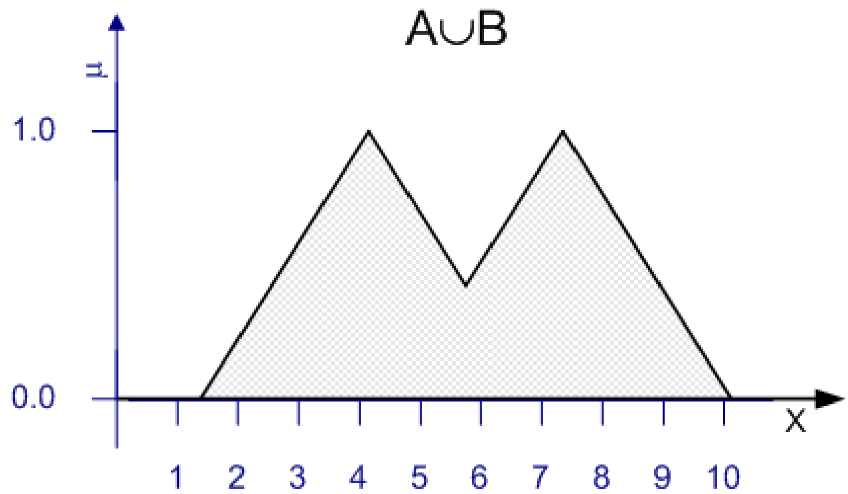


Figure 5.5.: $A \cup B$, the union of fuzzy sets A and B.

Definition 9 (Fuzzy union)**.** The union of fuzzy sets A and B is the fuzzy set A∪B with a membership grade for every $x \in X$ given by

$$(\mathsf{A} \cup \mathsf{B})\,(x) = \max[\mathsf{A}(x), \mathsf{B}(x)] \tag{5.14}$$

for all x in X. The maximum operator is represented by the symbol "$\vee$"so that $\max[a, b]$ can be written $a \vee b$ and the membership function for union can be written

$$(\mathsf{A} \cup \mathsf{B})\,(x) = \mathsf{A}(x) \vee \mathsf{B}(x)\,. \tag{5.15}$$

Example 33. Consider the fuzzy set A and B given in the following table:

x	A	B
a	1.0	0.3
b	0.7	0.6
c	0.4	0.9

(5.16)

The union of two fuzzy sets A and B is a fuzzy set that contains an element x of X at the maximum membership degree consistent with the membership grade of x in A and of x in B. Since c is in A to the degree 0.4 and c is in B to the degree 0.9 we know that c is in the union of the fuzzy sets A and B to the degree 0.9 since this is the maximum of the two membership values. The complete result is the fuzzy set A ∪ B with membership values as follows:

x	A ∪ B
a	1.0
b	0.7
c	0.9

(5.17)

5.5.4. Complement

The complement represents the notion of "not" in human language.. The complement of a set A is another set A^{c} that contains each element x in the universe X to the opposite degree that the original set contained x.

Definition 10 (Fuzzy complement)**.** The standard complement operator introduced by Zadeh is

$$\mathsf{A}^{\mathsf{c}}(x) = 1 - \mathsf{A}(x)\,. \tag{5.18}$$

This formula represents one of the most important differences between fuzzy set theory and standard set theory. In set theory it is always true that a set and its complement have nothing in common. In fuzzy set theory a set and its complement can be identical.

Example 34. Consider again the fuzzy set A of Example (29):

x	$\mathsf{A}(x)$
a	1.0
b	0.7
c	0.4

(5.19)

The complement of a fuzzy sets A contains an element x of X at one minus the membership grade of x in A. Since c is in A to the degree 0.4 we know that c is in the complement of the fuzzy sets A to the degree 0.6 since $0.6 = 1 - 0.4$. The complete result is the fuzzy set A^{c} with membership values as follows:

x	$\mathsf{A}^{\mathsf{c}}(x)$
a	0.0
b	0.3
c	0.6

(5.20)

and the membership grade of x in the intersection of A and A^{c}, $\mathsf{A} \cap \mathsf{A}^{\mathsf{c}}$, is

x	$\mathsf{A} \cap \mathsf{A}^{\mathsf{c}}$
a	0.0
b	0.3
c	0.4

(5.21)

This is a better picture of reality since, for instance, we could go to a movie and like and dislike it at the same time. People often hold contradictory feelings, opinions and evaluations about the same exact thing. And people often fluctuate in their evaluations. For instance suppose we order an unusual dish at a restaurant. Different diners may have conflicting opinions about the taste of a dish. An individual opinion may change, jumping from one side of a decision to the other with each new bite. A dish that seems too spicy initially can "grow on us."

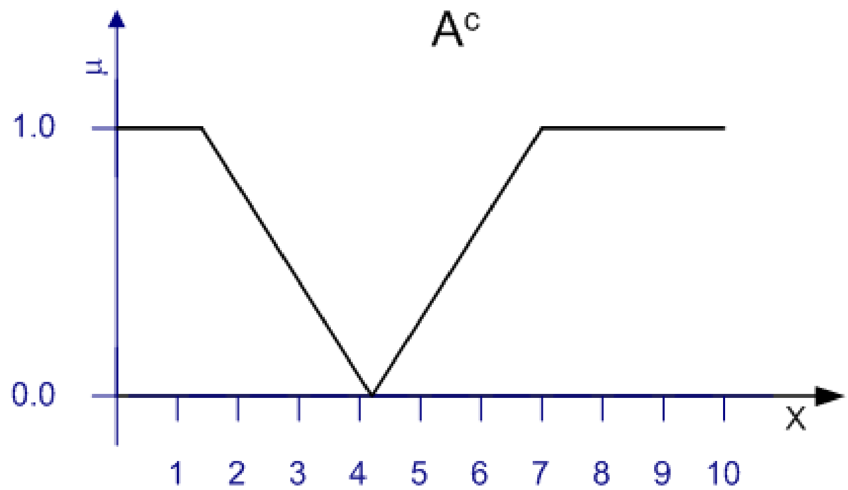


Figure 5.6.: A^c, the complement of A.

5.5.5. Set difference

If we examine a typical Venn diagram of two overlapping crisp sets, A and B then we see that there are essentially four pieces created by the intersection of A and B. There is the area of overlap, labeled $A \cap B$, and there is the area outside of both A and B. Then there are the lobes labeled $A \setminus B$ and $B \setminus A$ (sometimes this is written $A - B$ and $B - A$). The operation shown here is set subtraction, also called relative complement. In crisp sets, By $A - B$ we mean a set containing all the elements in A that are not in B. Thus, in crisp set theory, $A \setminus B = \{x \mid x \in A \, \text{and} \, x \notin B\}$. In fuzzy set theory this is not a useful definition. In fuzzy sets we assign the membership grade of $x \in X$ in $\mathsf{A} - \mathsf{B}$ as the difference of the membership grades, except that if the difference is negative, we set the membership grade to zero.

Definition 11 (Fuzzy set difference)**.** The difference of two fuzzy sets A and B is the fuzzy set $\mathsf{A} - \mathsf{B}$ with membership function

$$\mathsf{A} \setminus \mathsf{B}(x) = \max \left[\mathsf{A}(x) - \mathsf{B}(x), 0 \right] . \tag{5.22}$$

Example 35. Consider the fuzzy sets A and B given in the following table:

x	A	B
a	1.0	0.3
b	0.7	0.6
c	0.4	0.9

(5.23)

The difference of two fuzzy sets A and B is a fuzzy set that contains an element x of X at the difference of the membership grade of x in A and of x in B, or zero if this difference is negative. Since c is in A to the degree 0.4 and c is in B to the degree 0.9

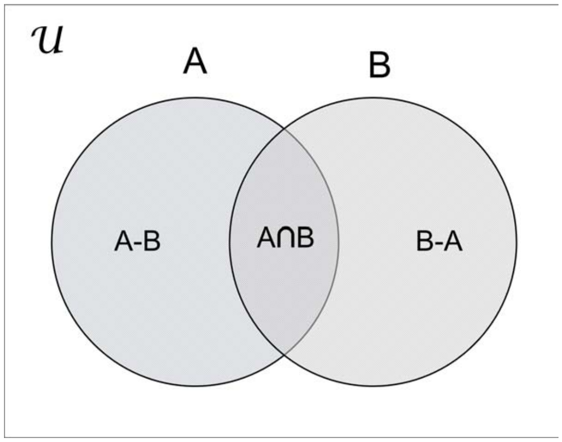


Figure 5.7.: $A - B$, the complement of A relative to B.

we know that c is in the union of the fuzzy sets A and B to the degree 0.0 since the difference is -0.5 which is replaced with zero. The complete result is the fuzzy set A − B with membership values as follows:

x	A − B
a	0.7
b	0.1
c	0.0

(5.24)

In crisp set theory, the simplest definition of set difference uses the formula $A - B \equiv A \cap B^{\mathsf{c}}$ however this equivalence is not true in fuzzy set theory. In general, the distinguishing difference between crisp and fuzzy set theory is the behavior of complementation. If a formula involves the complement then just because it is true in crisp set theory does not mean it is true in fuzzy. We illustrate this fact immediately.

Example 36. Consider the universe $X = \{a, bc\}$ with sets $A = \{a, b\}$ and $B = \{a, c\}$. Then $A - B = \{b\}$ and $A \cap B^{\mathsf{c}} = \{b\}$ which are identical. However consider the following table showing the fuzzy sets A, B. A − B, and $\mathsf{A} \cap \mathsf{B}^{\mathsf{c}}$:

x	A	B	A − B	$\mathsf{A} \cap \mathsf{B}^{\mathsf{c}}$
a	1.0	0.3	0.7	0.7
b	0.7	0.6	0.1	0.4
c	0.4	0.9	0.0	0.1

(5.25)

The final two columns produce fuzzy sets with markedly different membership grades. In fuzzy set theory $A - B \neq A \cap B^{\mathsf{c}}$.

5.6. Membership Grade Operations

5.6.1. Scalar cardinality

The scalar cardinality of a fuzzy set A is a count of the number of elements in A. Of course some elements are not completely in A so the scalar cardinality of A is not necessarily an integer.

Definition 12 (Scalar cardinality)**.** The scalar cardinality of A is the sum of the degree of membership of every element in A and is denoted $|\mathsf{A}|$.

$$|\mathsf{A}| = \sum_{x \in X} \mathsf{A}(x) \tag{5.26}$$

Example 37. The scalar cardinality of the fuzzy set A,

$$\begin{array}{c|c} x & \mathsf{A}(x) \\ \hline a & 1.0 \\ b & 0.7 \\ c & 0.4 \end{array} \tag{5.27}$$

from Example 29 is

$$|\mathsf{A}| = \sum_{x \in X} \mathsf{A}(x) \tag{5.28}$$

$$= \sum_{x \in \{a,b,c\}} \mathsf{A}(x) \tag{5.29}$$

$$= \mathsf{A}(a) + \mathsf{A}(b) + \mathsf{A}(c) \tag{5.30}$$

$$= 1.0 + 0.7 + 0.4 \tag{5.31}$$

$$= 2.1 \tag{5.32}$$

5.6.2. Alpha–cut or α–cut

A fuzzy set is a collection of objects with various degrees of membership. Often it is useful to consider those elements that have at least some minimal degree of membership α. This is liking asking who has a passing grade in a class, or a minimal height to ride on a roller coaster. We call this process an alpha–cut.

Definition 13. For every $\alpha \in [0, 1]$, a given fuzzy set A yields a crisp set A^α which contains those elements of the universe X who have membership grade in A of at least α:

$$\mathsf{A}^\alpha = \{x \in X \mid \mathsf{A}(x) \geq \alpha\}$$

We can not emphasize enough that an alpha–cut of a fuzzy set is not a fuzzy set, it is a crisp set.

Example 38. Consider the fuzzy set A given in the following table:

$$\begin{array}{c|l} x & \mathsf{A} \\ \hline a & 1.0 \\ b & 0.7 \\ c & 0.4 \end{array} \tag{5.33}$$

then the α–cut of A at $\alpha = 0.5$ is

$$\mathsf{A}^{0.5} = \{a, b\}$$

since $\mathsf{A}(a) = 1.0 \geq 0.5$ and $\mathsf{A}(b) = 0.7 \geq 0.5$ but $\mathsf{A}(c) = 0.4 \ngeq 0.5$

Since $\alpha < \beta$ implies $\mathsf{A}^\alpha \supseteq \mathsf{A}^\beta$ the set of all distinct α–cuts of any fuzzy set forms a nested sequence of crisp sets.

Remark 1. Klir champions the use of ${}^\alpha\mathsf{A}$ for the alpha cut since A^α could be interpreted as an exponent. While this is a salient point, the majority of books and papers in fuzzy set theory use A^α as the notation for an alpha–cut.

The *image set* of A, $I(\mathsf{A})$ (or, as it is sometimes denoted, $\mathsf{A}[X]$), is the image of the membership function μ_A. It consists of all values α in the unit interval such that $\mathsf{A}(x) = \alpha$ for some $x \in X$.

Definition 14 (Image set)**.** The image set of A is the set of all membership grades of $x \in X$ in A:

$$I(\mathsf{A}) = \{\mathsf{A}(x) \mid x \in X\} \ . \tag{5.34}$$

All the distinct alpha–cuts of a fuzzy set A are produced by values of alpha in the image set of A.

Example 39. Consider the fuzzy set A given in the following table:

$$\begin{array}{c|l} x & \mathsf{A} \\ \hline a & 1.0 \\ b & 0.7 \\ c & 0.4 \end{array} \tag{5.35}$$

The image set of A is $I(\mathsf{A}) = \{0.4, 0.7, 1.0\}$. The following three alpha—cuts are all the distinct sets that can be produced from A using alpha–cuts:

$$\mathsf{A}^{1.0} = \{a\} \tag{5.36}$$

$$\mathsf{A}^{0.7} = \{a, b\} \tag{5.37}$$

$$\mathsf{A}^{0.4} = \{a, b, c\} \tag{5.38}$$

Remark 2. Some books and papers will call an alpha–cut a cut–set and/or a level–set. Unfortunately Klir uses level set as the name for the image of the membership function.

In a certain sense, the three equations, 5.36—5.38, allow us to reconstruct the fuzzy set A from its pieces. We assign the membership grade of a to be the largest value of alpha such that a is in the alpha–cut associated with that value. In the discrete case this is simply the maximum value in the image set such the element is in the associated alpha–cut. As anyone who has taken calculus know, the infinity of the real numbers makes things harder to define easily.

Theorem 1 (Decomposition)**.** *Given an arbitrary fuzzy set* A*, it is uniquely represented by the associated sequence of its* α*–cuts via the formula*

$$\mathsf{A}(x) = \sup_{\alpha \in [0,1]} \alpha \cdot \chi_{\mathsf{A}^\alpha}(x) \tag{5.39}$$

where χ_{A^α} *denotes the characteristic function of the crisp set* A^α *and* $\sup$ *designates the supremum (or maximum, when the image set* $I(A)$ *is finite) which is the largest value the expression attains as* α *ranges over the unit interval.*

The characteristic function of an alpha–cut is either one or zero, depending on whether x is in the alpha–cut or not. When we multiply α by zero it will not be the maximum value. This formula, in essence, picks out the largest value of alpha such that $x \in \mathsf{A}^\alpha$.

Equation (5.39) is usually referred to as a *decomposition theorem* for fuzzy sets Zadeh (1971). It establishes an important connection between fuzzy sets and crisp sets. This connection provides us with a criterion for generalizing properties of crisp sets into their fuzzy counterparts. To generalize to fuzzy set theory a property established for crisp set theory, this property should be preserved (in the crisp sense) in all α-cuts of the fuzzy sets involved. For example, all α-cuts of a convex fuzzy sets should be convex crisp sets. Each α-cut of a properly defined fuzzy equivalence relation (or fuzzy compatibility relation, fuzzy ordering relation, etc.) should be an equivalence relation (or compatibility relation, ordering relation, etc., respectively) in the classical sense. A property of a fuzzy set that is preserved in all its α-cuts is called a *cutworthy property*.

Example 40. Let $X = \{0, 1, 2, \dots, 29, 30\}$

$$\mathsf{B}(x) = \frac{30x - x^2}{225} \tag{5.40}$$

Then $\mathsf{B}(0) = 0$, $\mathsf{B}(5) = 0.56$, $\mathsf{B}(15) = 1$, $\mathsf{B}(20) = 0.89$, etc.

The α cut of B at $\alpha = 0.7$ is a crisp set containing all those elements $x \in X$ whose membership grade is greater than or equal to 0.7, that is $\mathsf{B}^\alpha = \{7, 8, 9, \dots, 22, 23\}$.

Most proofs in this book are located in the appendices. However the following theorem is important enough to include in place. It shows that a fuzzy set is completely characterized by its alpha–cuts.

Theorem 2. *Two fuzzy sets are equal if and only if all their corresponding* α *cuts are equal.*

Proof. We are trying to prove that :

$$\forall \alpha \, : \, \mathsf{A}^\alpha = \mathsf{B}^\alpha \longleftrightarrow \mathsf{A} = \mathsf{B}. \tag{5.41}$$

First let us show that : $\mathsf{A} = \mathsf{B} \rightarrow \mathsf{A}^\alpha = \mathsf{B}^\alpha$. When we say that $\mathsf{A} = \mathsf{B}$ we mean that their membership functions are identical for every $x \in X$. First $\mathsf{A}^\alpha = \{x \mid \mathsf{A}(x) \geq \alpha\}$ and $\mathsf{B}^\alpha = \{x \mid B(x) \geq \alpha\}$ but $\mathsf{A}(x) = \mathsf{B}(x)$ for all x so $\mathsf{A}^\alpha = \mathsf{B}^\alpha$ for all α.

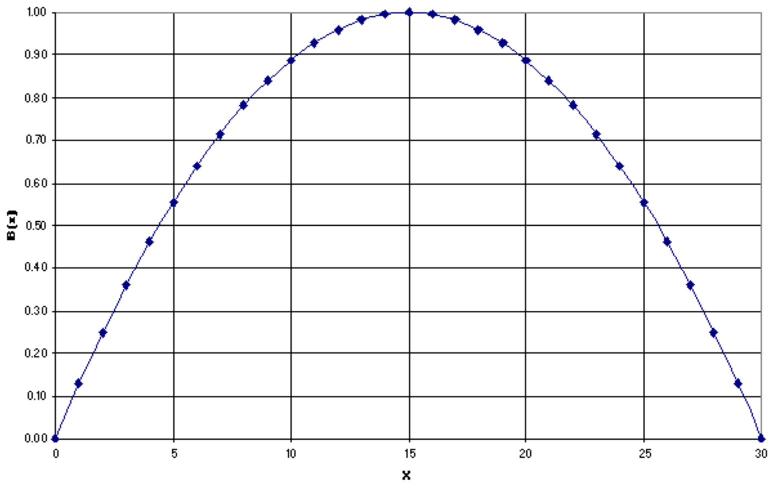


Figure 5.8.: The membership grade of the x values $0, 1, 2, \ldots, 30$ in the fuzzy set B.

To show the reverse, that : $\forall\alpha\ \mathsf{A}^\alpha = \mathsf{B}^\alpha \rightarrow \mathsf{A} = \mathsf{B}$ suppose that $\forall\alpha\ \mathsf{A}^\alpha = \mathsf{B}^\alpha$ but $\mathsf{A} \neq \mathsf{B}$. But $\mathsf{A} \neq \mathsf{B}$ if and only if there exists some $y \in X$ such that $\mathsf{A}(y) \neq \mathsf{B}(y)$. Without loss of generality assume that $\mathsf{A}(y) < \mathsf{B}(y)$ and let $\gamma = \mathsf{B}(y)$. It must be that $y \notin \mathsf{A}^\gamma$ but $y \in \mathsf{B}^\gamma$. Then the α-cuts of A and B are not identical, and this is a contradiction. □

The *strong alpha cut*, $\mathsf{A}^{\alpha+}$, is defined to be all those elements that have membership strictly greater than alpha,

$$\mathsf{A}^{\alpha+} = \{x \mid \mathsf{A}(x) > \alpha\}. \tag{5.42}$$

A couple of alpha–cuts are special. The strong alpha–cut at zero contains all the elements of a fuzzy set with positive membership grade. This is called the support of the fuzzy set. The alpha–cut at one of a fuzzy set contains all those elements of a fuzzy set with complete membership. This is called the core of a fuzzy set.

Definition 15 (Support)**.** The *support* of a fuzzy set is the strong alpha-cut at zero:

$$S(\mathsf{A}) = \mathsf{A}^{0+} \tag{5.43}$$

$$= \{x \mid \mathsf{A}(x) > 0\}\ . \tag{5.44}$$

Hence it is that subset of the domain which has positive membership grade in a fuzzy set.

Example 41. The support of A (whose membership function is given in 5.3) is $\{a, b, c\}$ and the support of B (whose membership function is given in 5.40) is $\{2, 3, 4, \ldots, 28, 29\}$.

Definition 16 (Core)**.** The *core*, peak or mode of a fuzzy set is the alpha-cut at one:

$$C(\mathsf{A}) = \mathsf{A}^1. \tag{5.45}$$

$$= \{x \mid \mathsf{A}(x) = 1\}\ . \tag{5.46}$$

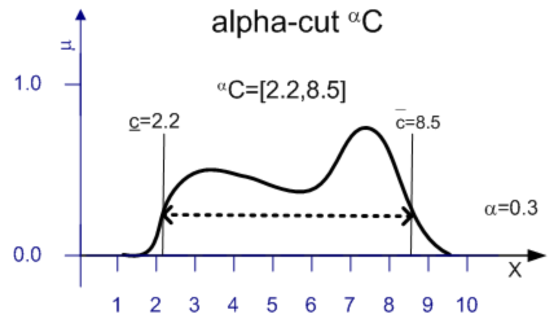


Figure 5.9.: Alpha-cut of a fuzzy set C an a continuous domain X.

Example 42. The core of A is $\{a\}$ and the core of B is $\{15\}$.

In the image set of a fuzzy set two membership grades are especially important, the smallest and the largest. The smallest is called the plinth of the fuzzy set and the largest is called the height of a fuzzy set.

Definition 17 (Height)**.** The *height* of a fuzzy set A, $h(\mathsf{A})$, is the maximum or supreme membership grade achieved by $\mathsf{A}(x)$ over the domain of X :

$$h(\mathsf{A}) = \sup_{x \in X}(x) \ . \tag{5.47}$$

If the height of a fuzzy set is equal to 1.0 then we call that fuzzy set *normal*.

Example 43. The height of A is 1.0 and the height of B is 1.0. Thus they are both normal fuzzy sets.

Definition 18 (Plinth)**.** The *plinth* of a fuzzy set A, $p(\mathsf{A})$, is the minimum, or infimum, membership grade achieved by $\mathsf{A}(x)$ over the domain of X :

$$p(\mathsf{A}) = \inf_{x \in X}(x) \ . \tag{5.48}$$

Example 44. The plinth of A is 0.4 and the plinth of B is 0.0.

Definition 19 (Normal)**.** A fuzzy set is normal if there is some $x \in X$ such that its membership grade in the fuzzy set is 1.0.

5.7. Possibility Theory

5.7.1. Possibility distributions

Lotfi Zadeh, the architect of fuzzy set theory, is also responsible for the creation of possibility theory which he introduced in the paper "*Fuzzy sets as a basis for a theory of possibility,*" Zadeh (1978) Zadeh interpreted a fuzzy set B, representing a proposition **B**, as a fuzzy restriction on the domain set X. He then set the *possibility of* **B** *being* X is equal to the fuzzy restriction grade $\mathsf{B}(X)$. In many ways the literature of possibility theory is also the literature of fuzzy sets. Possibility theory is special in two ways. First, it is assumed that there is first element that is totally possibly **B**. Second it is assumed that each subsequent element has a smaller degree of possibility of being **B.**

Formally a possibility distribution, $q_,$ is a mapping from some universal set X into the unit interval,

$$\mathsf{q} : X \to \mathbb{I}.$$

If $|X| = n$ then $\mathsf{q}_i = \mathsf{q}(x_i)$, $\mathbb{I} \in \mathbb{N}_n$, will designate an arbitrary element in the ordered n—tuple

$$\mathbf{q} = \langle \mathsf{q}_1, \mathsf{q}_2, \ldots, \mathsf{q}_i, \ldots, \mathsf{q}_n \rangle$$

with

$$1 = \mathsf{q}_1 \geq \mathsf{q}_2 \geq \ldots \geq \mathsf{q}_i \geq \ldots \geq \mathsf{q}_n \geq 0$$

Just as probabilities can be derived from a finite set of frequencies by dividing by their sum, possibilities can be derived from any finite set of non-negative numbers by dividing by the maximum and re-ordering the universal set.

Remark 3. Many applications are concerned with using the paradigms of possibility theory to process real world data. Data does not come pre-sorted and when dealing with data derived from an image the researcher is not at liberty to re-arrange the image, a pixel's original position is an important piece of information.

Example 45. If a fuzzy set B (for "blue") is defined with membership values for five clothing items labeled $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$, $\dot{x}_4$, $\dot{x}_5$ then it is feasible to interpret the membership grade of the item tagged with the label $\dot{x}_3$ as a gauge of the possibility of $\dot{x}_3$ being blue. This value can range from zero, or impossible, to one, or surely possible. Numerical data, when normalized and treated as a possibility distribution, has two essential characteristics, relative magnitude and absolute order. If $\dot{x}_i$ is again an object of clothing and B is a fuzzy set representing the verbal concept blue then by some process one might calculate the evaluations $\mathsf{B}(\dot{x}_1) = 0.3$, $\mathsf{B}(\dot{x}_2) = 0.2$, $\mathsf{B}(\dot{x}_3) = 0.4$, $\mathsf{B}(\dot{x}_4) = 1.0$ and $\mathsf{B}(\dot{x}_5) = 0.7$ for the membership grades of the elements $\dot{x}_1$, $\dot{x}_2$, $\dot{x}_3$, $\dot{x}_4$, and $\dot{x}_5$. These values might represent the subjective opinion of an expert on the quality or degree of blueness, or they may represent an absolute measurement of the amount of blue light registered by some instrumentation. In either case the values must be non-negative and one evaluation must be one for the data to be considered as a possibility distribution.

If all the evaluations are positive but none are equal to one the fuzzy set can be normalized. This normalization is almost always done by dividing through all the blueness values by the maximum value. As shall be seen in the following sections, this is not the only method of deriving possibility distributions from fuzzy sets. Raw data can provide a possibility distribution if there is some (monotone) transformation of the data values into the unit interval with at least one value mapped to a possibility grade of 1.

Since, in Zadeh's interpretation, a possibility distribution is induced by a fuzzy set we have reorder the universal set based on fuzzy membership grades. This gives the permutation $x_1 = \dot{x}_4$, $x_1 = \dot{x}_5$, $x_3 = \dot{x}_3$, $x_4 = \dot{x}_1$, $x_5 = \dot{x}_2$ and the possibility distribution

$\mathbf{q} = \langle 1.0, 0.7, 0.4, 0.3, 0.2 \rangle$.

$$\begin{aligned}
\mathsf{q}(x_1) &= B(\dot{x}_4) = 1.0, \\
\mathsf{q}(x_2) &= B(\dot{x}_5) = 0.7, \\
\mathsf{q}(x_3) &= B(\dot{x}_3) = 0.4, \\
\mathsf{q}(x_4) &= B(\dot{x}_1) = 0.3 \text{ and} \\
\mathsf{q}(x_5) &= B(\dot{x}_2) = 0.2 \,.
\end{aligned}$$

5.7.2. Possibility measures

A possibility distribution is a function from a set X into the unit interval with some element of the universe having membership grade one. This distribution induces a possibility measure upon the subsets of X via the formula

$$\mathsf{Q}(A) = \max_{x \in A} \mathsf{q}(x) \tag{5.49}$$

if X is finite. If X is infinite we must use $\sup$ for $\max$ in Eq. (5.49). The possibility measure Q is obviously a function from the power set of X into the interval $[0,1]$, that is, $\mathsf{Q} : \mathcal{P}(X) \to \mathbb{U}$.

A possibility measure can be defined directly without recourse to a possibility distribution. A possibility measure is a mapping that is zero on the empty set,

$$\mathsf{Q}(\emptyset) = 0$$

one on the universal set,

$$\mathsf{Q}(X) = 1$$

and such that the possibility measure of the union of two sets is equal to the maximum of the possibility measures of the two component sets,

$$\mathsf{Q}(A \cap B) = \max[\mathsf{Q}(A), \mathsf{Q}(B)]$$

Given a possibility measure Q it is easy to construct its associated possibility distribution from the formula

$$\mathsf{q}(x_i) = v(\{x_i\})$$

Eq. (5.49) gives the formula for constructing a possibility measure from its associated possibility distribution.

Furthermore, it is easy to show that

$$\mathsf{Q}(A) + \mathsf{Q}(A^{\mathsf{c}}) \geq 1 \,, \tag{5.50}$$

$$\max[\mathsf{Q}(A), \mathsf{Q}(A^{\mathsf{c}})] = 1 \,. \tag{5.51}$$

Ordered possibility distributions of the same length can be partially ordered in the following way: given two possibility distributions

$$\mathbf{q}_i = \langle \mathsf{q}_{i1}, \mathsf{q}_{i2}, \ldots, \mathsf{q}_{in} \rangle$$

and

$$\mathbf{q}_j = \langle \mathsf{q}_{j1}, \mathsf{q}_{j2}, \ldots, \mathsf{q}_{jn} \rangle$$

we define

$$\mathbf{q}_i \leq \mathbf{q}_j \text{ if and only if } \mathsf{q}_{ik} \leq \mathsf{q}_{jk}$$

for all $k = 1, 2, ..., n$. This partial ordering forms a lattice whose join, $\vee$, and meet, $\wedge$, are defined, respectively, as

$$\mathbf{q}_i \vee \mathbf{q}_j = \langle \max[\mathsf{q}_{i1}, \mathsf{q}_{j1}], \max[\mathsf{q}_{i2}, \mathsf{q}_{j2}], \ldots, \max[\mathsf{q}_{in}, \mathsf{q}_{jn}] \rangle$$

and

$$\mathbf{q}_i \wedge \mathbf{q}_j = \langle \min[\mathsf{q}_{i1}, \mathsf{q}_{j1}], \min[\mathsf{q}_{i2}, \mathsf{q}_{j2}], \ldots, \min[\mathsf{q}_{in}, \mathsf{q}_{jn}] \rangle$$

for all pairs of possibility distributions of the same length n. The smallest possibility distribution is $\langle 1, 0, ..., 0 \rangle$, the greatest one is $\langle 1, 1, ..., 1 \rangle$; their counterparts expressing the basic probability assignment are $\langle 1, 0, ..., 0 \rangle$ and $\langle 0, 0, ..., 0, 1 \rangle$, respectively. We can see from this that in possibility theory the expression of full certainty and, contrary to probability theory, also the expression of total ignorance are exactly the same as in evidence theory.

Finally we note that the literature on possibility theory uses either P, Π, or Q to represent a possibility measure and p, π, or q to represent the corresponding possibility distribution. This book will always use Q and q or the ordered possibility measure and distribution.

5.8. Advanced Types of Fuzzy Sets

For some applications, it is useful to define fuzzy sets in terms of more general forms of membership grade functions. An important form is

$$\mathsf{A} : X \to L, \tag{5.52}$$

where L denotes a lattice. Fuzzy sets defined by functions of this form are called *L-fuzzy sets.* Lattice L may, for example, consist of a class of closed intervals in $[0, 1]$. Membership degrees are in this case defined imprecisely, by closed sub-intervals of $[0, 1]$. Fuzzy sets with this property are called *interval–valued fuzzy sets*. When L is a class of fuzzy numbers defined on $[0, 1]$, we obtain *fuzzy sets of type-2*. We will have little use of these notions in this book.

5.9. Notes

Most of Zadeh's papers are collected in Klir and Yuan (1996). Good books on fuzzy set theory include Klir and Yuan (1995), Dubois and Prade (1980a), and Klir and Folger (1988). For general applications the best books, other than the one you are holding, are Jang et al. (1997), Kandel (1986), Zimmermann (1996), Kaufmann (1975), and Kaufmann and Gupta (1985).

5.10. Homework

Let us define the universal sets

$$X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}, \tag{5.53}$$

$$Y = \{a, b, c\}, \tag{5.54}$$

and

$$Z = \{\alpha, \beta, \gamma, \delta\}. \tag{5.55}$$

Next we present the membership functions of the fuzzy set A, B, C, D,E, and F. The membership functions for E and F are presented a list of ordered pairs with the interpretation of $\langle 1, 0.2\rangle$ being that the membership grade of 1 in E is 0.2, or $\mathsf{E}(1) = 0.2$. If an element of the domain set is not listed the interpretation is that the membership grade is zero. Thus the membership grade of 7, which is in X, in E is 0.0. Fuzzy sets defined on X use the variable x and likewise we use y and z for arbitrary elements of Y and Z.

$$\begin{aligned}
\mathsf{A}(x) &= \frac{x}{10} \\
\mathsf{B}(x) &= \frac{|x-5|}{5} \\
\mathsf{C}(x) &= \begin{cases} \frac{9-|2x-9|}{8} & 1 \leq x \leq 8 \\ 0 & otherwise \end{cases} \\
\mathsf{D}(x) &= 0.8 \\
\mathsf{E}(x) &= \{\langle 1, 0.2\rangle, \langle 2, 0.6\rangle, \langle 3, 0.4\rangle\} \\
\mathsf{F}(y) &= \{\langle a, 0.3\rangle, \langle b, 0.7\rangle, \langle c, 0.9\rangle\} \\
\mathsf{G}(z) &= \{\langle \alpha, 0.5\rangle, \langle \beta, 0.8\rangle, \langle \gamma, 0.35\rangle\}
\end{aligned}$$

Given the above information, try to answer the following questions.

1. For the sets, A through G, what universe do they belong to?
2. What is A^c ?
3. What is B^c ?
4. What is C^c ?
5. What is G^c ?
6. What is $\mathsf{A} \cup \mathsf{B}$?
7. What is $\mathsf{A} \cup \mathsf{C}$?
8. What is $\mathsf{B} \cup \mathsf{C}$?
9. What is $\mathsf{A} \cup \mathsf{B}^c$?
10. What is $\mathsf{A} \cup \mathsf{C}^c$?

11. What is $\mathsf{B} \cup \mathsf{C}^c$?
12. What is $\mathsf{A} \cup \mathsf{A}^c$?
13. What is $\mathsf{A} \cap \mathsf{B}$?
14. What is $\mathsf{A} \cap \mathsf{C}$?
15. What is $\mathsf{B} \cap \mathsf{C}$?
16. What is $\mathsf{A} \cap \mathsf{B}^c$?
17. What is $\mathsf{A} \cap \mathsf{C}^c$?
18. What is $\mathsf{B} \cap \mathsf{C}^c$?
19. What is $\mathsf{A} \cap \mathsf{A}^c$?
20. What is $\mathsf{F} \cap \mathsf{G}$?
21. What is $\mathsf{A} \setminus \mathsf{B}$ and $\mathsf{B} \setminus \mathsf{A}$.
22. What is $\mathsf{A} \setminus \mathsf{C}$ and $\mathsf{C} \setminus \mathsf{A}$.
23. What is $\mathsf{B} \setminus \mathsf{C}$ and $\mathsf{C} \setminus \mathsf{B}$.
24. How would you go about defining $\mathsf{A} \times \mathsf{B}$?
25. What is the image set, core, support, height, and plinth of A?
26. What is the image set, core, support, height, and plinth of B?
27. What is the image set, core, support, height, and plinth of C?
28. What is the image set, core, support, height, and plinth of D?
29. What is the image set, core, support, height, and plinth of E?
30. What is the image set, core, support, height, and plinth of G?
31. What is A^α and $\mathsf{A}^{\alpha+}$
    a) if $\alpha = 0.0$?
    b) if $\alpha = 0.2$?
    c) if $\alpha = 0.4$?
    d) if $\alpha = 0.6$?
    e) if $\alpha = 0.8$?
    f) if $\alpha = 1.0$?
32. What is B^α and $\mathsf{B}^{\alpha+}$
    a) if $\alpha = 0.0$?
    b) if $\alpha = 0.2$?
    c) if $\alpha = 0.4$?

d) if $\alpha = 0.6$?
e) if $\alpha = 0.8$?
f) if $\alpha = 1.0$?

33. What is C^{α} and $\mathsf{C}^{\alpha+}$
    a) if $\alpha = 0.0$?
    b) if $\alpha = 0.2$?
    c) if $\alpha = 0.4$?
    d) if $\alpha = 0.6$?
    e) if $\alpha = 0.8$?
    f) if $\alpha = 1.0$?
34. Who is Lotfi Zadeh?
35. Why fuzzy sets?

6. Fuzzy operators

6.1. Introduction

Many famous logicians of 20th century introduced multi-valued logics. The Polish logician and philosopher Jan Łukasiewicz (1920) introduced a three valued logic, using true, false, and a third value to represent the unknown. The stimulus for this creation was Aristotle's paradox of the sea battle, "There will be a sea battle December 12, 2012" is neither true nor false, it is unknown. Even worse, the day after tomorrow, we will know whether or not the statement about the sea battle is true or not. Since something that is true must always have been true, and always will be true, the statement must have a truth value that annot be determined. But then we must ask ourselves if any truth value can be determined?

Another famous mathematician, the Emil L. Post (1921) also introduced n-valued logics, where $n \geq 2$. In 1932 Jan Łukasiewicz and Alfred Tarski introduced another, slightly different n-valued logic. The most famous logician of the 20th century, Kurt Gödel, in 1932 defined a system of multi-valued logics.

Not surprisingly, they did not all introduce identical systems, see 3.6. For example, they did not all agree on how to calculated logical and with their new fractional values. Thus it is not surprising that researchers in fuzzy set theory have introduced more than one way to model and.

Fuzzy set theory is a new discipline, and that has some advantages and some drawbacks. The drawbacks include multiple names and notations for identical objects. The advantages are an appreciation for innovation and originality.

This chapter will focus on operators. Operators take one, two, or more fuzzy sets and produce another fuzzy set. Most of the operators will take two fuzzy sets and produce a third fuzzy set as a result, these type of operators are called binary operators. We have already seen many of these operators in the earlier chapters of this book. The $\min$ operator that is used for the standard intersection is one example. If the fuzzy sets are fuzzy numbers then addition is another binary operator that takes two fuzzy numbers and produces their sum. An operator that takes only one argument is called a unary operator. The standard fuzzy complement is a unary operator.

Let us go back to pure set theory. If $x \in A$ and $x \in B$ then it is in $A \cap B$, otherwise x is not in the intersection of A and B. If we write a table in terms of characteristic functions for the sets A, B, and the resulting characteristic function of $A \cap B$ the intersection operator we get:

Note that the standard fuzzy operator for intersection, $\min$, will produce the same results in the third column given the membership grades in the first two columns. However *multiplication* will also produce these results. The question then is, if we replace the $\min$ operator with multiplication do we produce a consistent theory. If we do in fact replace $\min$ with multiplication, then what would be the proper replacement for the $\max$ operator? And how about the complement operator? Or should we have asked these questions another way round, and replaced $\max$ first?

In this chapter we will limit ourselves to variations on the basic set operators intersection, union and complementation. In addition we will take a brief look at aggregation operators, operators that do not have a direct correspondence in classical set theory.

6.2. Fuzzy Operators

A single fuzzy set can be operated on by the application of a *fuzzy complement*. Two fuzzy sets can be combined by one of three types of aggregating operations: *fuzzy intersection*, *unions*, and *averaging operations*. Multiple fuzzy sets can be combined with *aggregation* operators. Operations of each of these types are not unique. The whole scope of operations of each type can be conveniently captured by a class of functions distinguished from one another by distinct values of a parameter taken from a specific range of values. The choice of a particular operation is determined by the purpose for which it is used.

6.2.1. Standard fuzzy set operators

By far, the most important and common fuzzy complement, intersection and union operations are those defined by the formulas

$$\mathsf{A}^{\mathsf{c}}(x) = 1 - \mathsf{A}(x), \tag{6.1}$$
$$(\mathsf{A} \cap \mathsf{B})\,(x) = \min[\mathsf{A}(x), \mathsf{B}(x)] = \mathsf{A}(x) \vee \mathsf{B}(x), \tag{6.2}$$
$$(\mathsf{A} \cup \mathsf{B})\,(x) = \max[\mathsf{A}(x), \mathsf{b}(x)] = \mathsf{A}(x) \wedge \mathsf{B}(x). \tag{6.3}$$

Axiomatic characterization of these operations, which are usually referred to as *standard fuzzy operations*, was investigated by Bellman and Giertz (1973). Any property generalized from classical set theory into the domain of fuzzy set theory that is preserved in all α-cuts for $\alpha \in (0, 1]$ is called *cutworthy*. An arbitrary binary operator $\star$ is *idempotent* if $a \star a = a$. The minimum operation is the only fuzzy intersection that is idempotent and cutworthy; similarly, the maximum operation is the only union that is idempotent and cutworthy. No fuzzy complement is cutworthy.

$\chi_A(x)$	$\chi_B(x)$	$\chi_{A\cap B}(x)$
0	0	0
0	1	0
1	0	0
1	1	1

Table 6.1.: A table for the intersection operator.

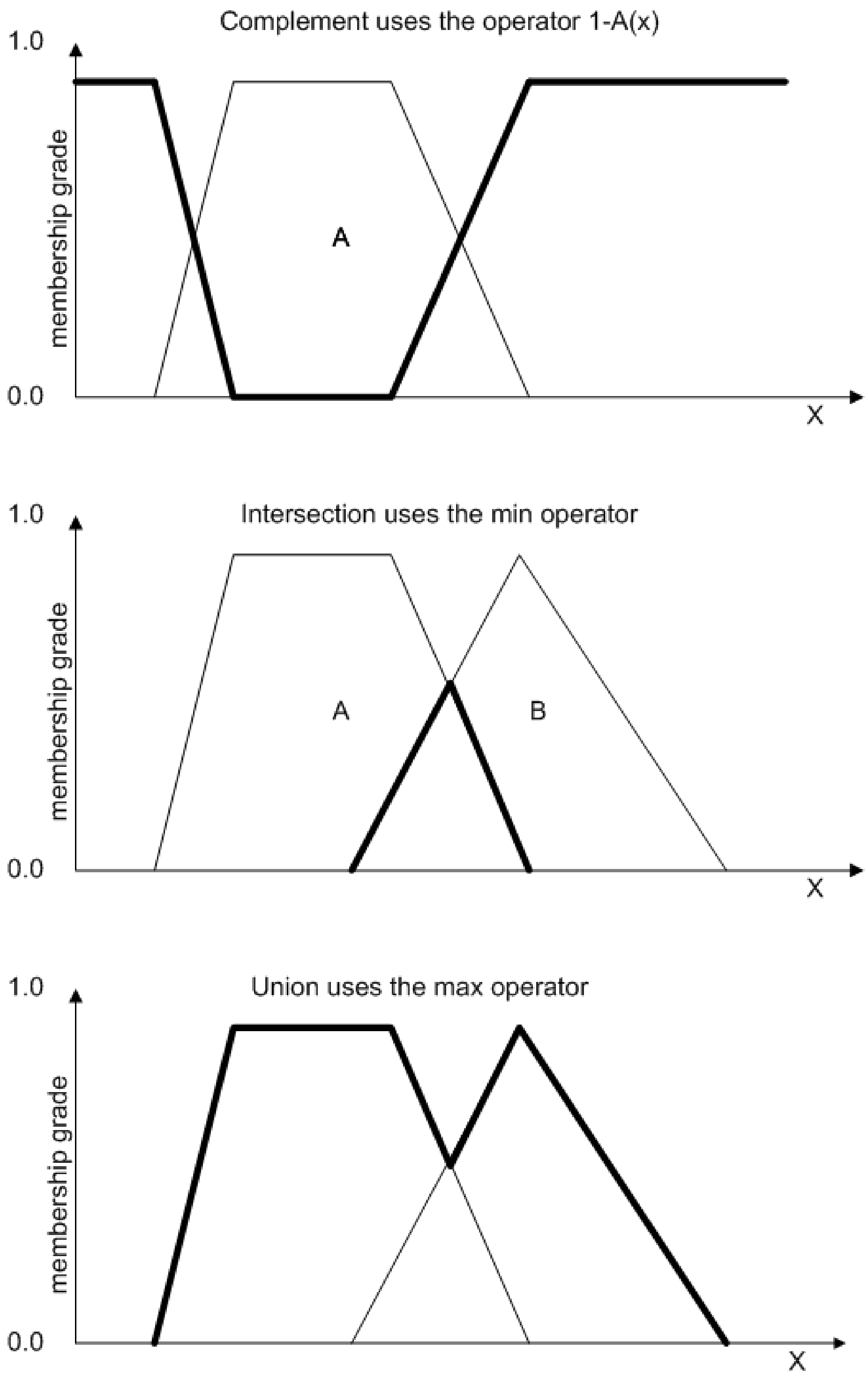


Figure 6.1.: Illustrating the standard fuzzy operators.

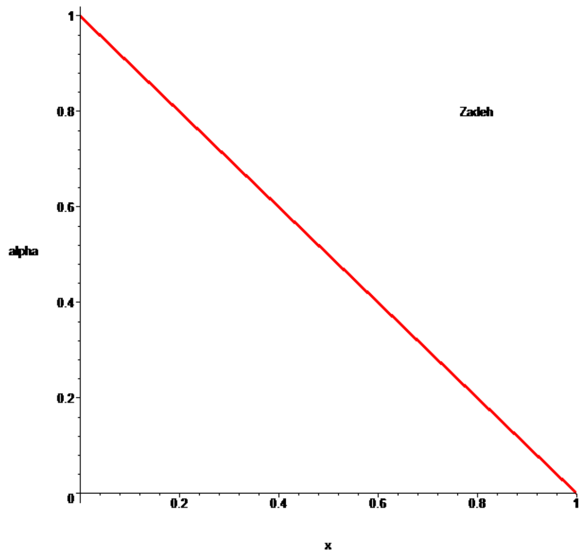


Figure 6.2.: Standard complement function.

6.3. Fuzzy Complement

Before we introduce alternatives to Zadeh's original operator for complements we should determine what kinds of unary operators fulfill our notions of what the opposite of a fuzzy set should be. One of the first conditions is that when a fuzzy set has all its membership values in the set $\{0, 1\}$ then the fuzzy set is essentially a crisp set. In this case, the complement should be the same as the complement of a crisp set, so that zero becomes one and vice versa. Second of all the membership degree of an element x in the complement of a fuzzy set $\mathsf{A}(x)$ should be a local phenomena, that is, it should depend only on the degree of membership of x in the original fuzzy set $\mathsf{A}(x)$. Finally the more x is in $\mathsf{A}(x)$ the less x is in the complement of $\mathsf{A}(x)$. These ideas are gathers together in the following requirements.

An arbitrary complement operators, $\mathsf{c} : [0, 1] \to [0, 1]$, must satisfy the following three axioms:

(c1) Membership dependency — The membership grade of x in the complement of A depends only on the membership grade of x in A.

(c2) Boundary condition — $\mathsf{c}(0) = 1$ and $\mathsf{c}(1) = 0$, that is c behaves as the ordinary complement for crisp sets.

(c3) Monotonicity — For all $a, b \in [0, 1]$, if $a < b$, then $\mathsf{c}(a) \geq \mathsf{c}(b)$, that is c is monotonic nonincreasing.

Two additional axioms, which are usually considered desirable, constrain the large family of functions that would be permitted by the above three axioms; they are:

(c4) Continuity — c is continuous.

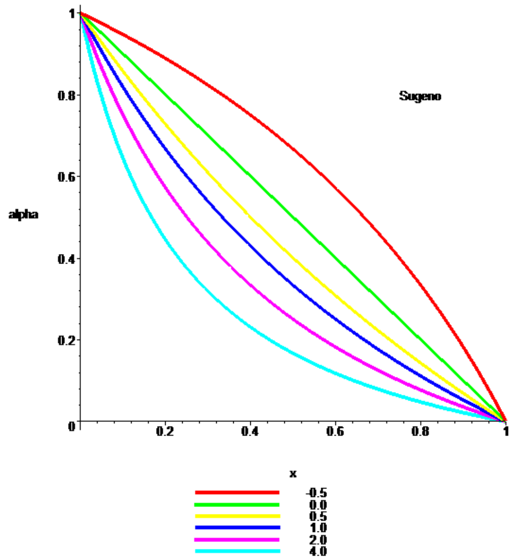


Figure 6.3.: Sugeno complement functions.

(c5) Involution — c is involutive, that is $\mathsf{c}(\mathsf{c}(a)) = a$.

The standard fuzzy complement function is:

$$\mathsf{c}(a) = 1 - a$$

and it is graphed in Fig. (6.2).

Some of the functions that conform to these five axioms besides the standard fuzzy complement are in the Sugeno class of fuzzy complements defined for all $a \in [0,1]$ by

$$\mathsf{c}[\lambda](a) = \frac{1-a}{1+\lambda a}, \tag{6.4}$$

with $\lambda \in (-1, \infty)$. The curves generated by the Sugeno complement for various values of λ are illustrated in Fig (6.3).

The Yager class of fuzzy complements defined for all $a \in [0,1]$ by

$$\mathsf{c}[w](a) = (1 - a^w)^{1/w}, \tag{6.5}$$

with $w \in (0, \infty)$. . The curves generated by the Yager complement for various values of w are illustrated in Fig (6.4).

Observe that the standard fuzzy complement, $\mathsf{c}(a) = 1 - a$, is obtained as the Sugeno complement at zero, $\mathsf{c}[\lambda = 0]$ and as the Yager complement at one, $\mathsf{c}[w = 1]$.

An example of fuzzy complements that conforms to **(c1)–(c3)** but not to **(c4)** and **(c5)** are the threshold fuzzy complements

$$\mathsf{c}[t](a) = \begin{cases} 1 & \text{when } a \in [0,t] \\ 0 & \text{when } a \in (t,1] \end{cases}, \tag{6.6}$$

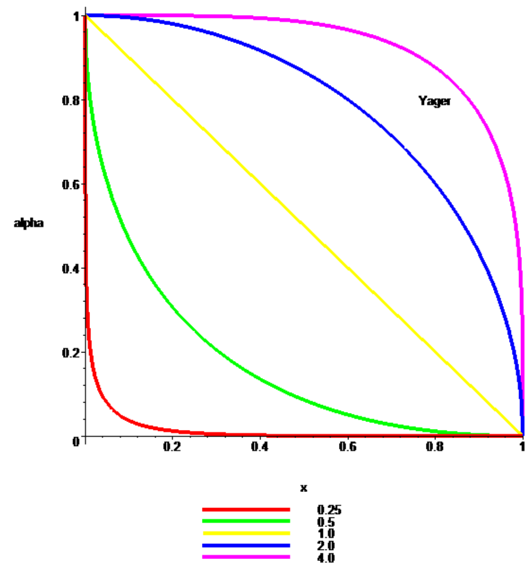


Figure 6.4.: Yager complement functions.

with $t \in [0, 1]$.

Subsequently, we shall write A^{c} for an arbitrary complement of the fuzzy set A; its membership function is $A^{\mathsf{c}}(x) = \mathsf{c}(A(x))$.

An equilibrium, e_{c}, of a fuzzy complement c, if it exists, is a number in $[0, 1]$ for which $\mathsf{c}(e_{\mathsf{c}}) = e_{\mathsf{c}}$. Every fuzzy complement has at most one fuzzy equilibrium and if a fuzzy complement is continuous (i.e., if it satisfies axioms **(c1)–(c4)**), the existence of its equilibrium is guaranteed Klir and Yuan (1996). For example, the equilibrium of fuzzy complements in the Yager class (6.5) are

$$e_{\mathsf{c}}[w] = 0.5^{\frac{1}{w}} \tag{6.7}$$

for each $w \in (0, \infty)$.

6.4. Fuzzy Set Intersections

In the previous section we saw that complementation can be based on a function c that manipulates membership values. Using the same line of reasoning, intersection of fuzzy sets A and B can be based upon a function t that that takes two arguments. The function t takes the membership grade of x in the fuzzy set A and the membership grade of x in the fuzzy set B and returns the membership grade of x in the fuzzy set $\mathsf{A} \cap \mathsf{B}$.

As was done for the complement, we will propose a reasonable axiomatic skeleton. The axioms will capture the basic notions of how intersection must operate on fuzzy sets. First off, fuzzy set intersection should work the same as intersection whenever the fuzzy sets have membership grades restricted to the set $\{0, 1\}$, as previously de-

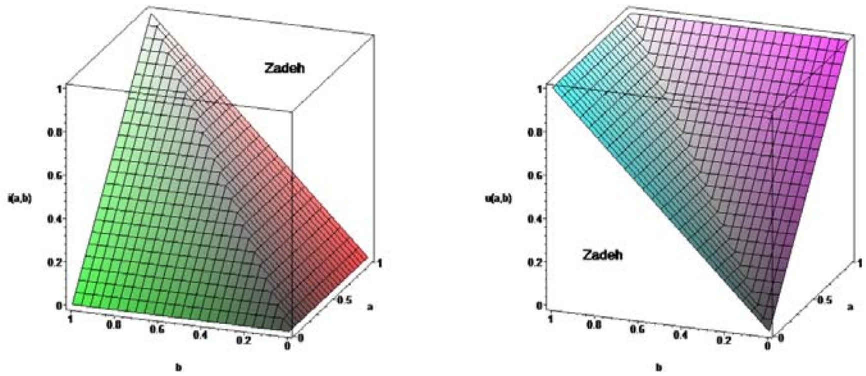


Figure 6.5.: Graphs of the max and min functions.

scribed in Table (6.1). Since $\mathsf{A} \cap \mathsf{B} = \mathsf{B} \cap \mathsf{A}$ the function should be commutative. Since $(\mathsf{A} \cap \mathsf{B}) \cap \mathsf{C} = \mathsf{A} \cap (\mathsf{B} \cap \mathsf{C})$ the function should be associative. Lastly the more x is in fuzzy set A and B the more it should be in $\mathsf{A} \cap \mathsf{B}$.

The intersection of two fuzzy sets must be a function that maps pairs of numbers in the unit interval into the unit interval $[0, 1]$, $\mathsf{t} : [0, 1] \times [0, 1] \to [0, 1]$. It is now well established that *triangular norms* or t-norms do possess all properties that are intuitively associated with fuzzy intersections. These functions are, for all $a, b, c, d \in [0, 1]$, characterized by he following axioms:

(i1) Boundary condition — $\mathsf{t}(1, a) = a$.

(i2) Monotonicity — $\mathsf{t}(a, b) \leq \mathsf{t}(c, d)$ whenever $a \leq c$ and $b \leq d$.

(i3) Commutativity — $\mathsf{t}(a, b) = \mathsf{t}(b, a)$.

(i4) Associativity — $\mathsf{t}(a, \mathsf{t}(b, c)) = \mathsf{t}(\mathsf{t}(a, b), c)$.

It turns out that functions that obey these four rules had been extensively studied in the literature of probability, long before the creation of fuzzy set theory. Functions that possess properties (**i1**) through (**i4**) were given the name t-norm Menger (1942), short for triangular-norm.

Axiomatic skeletons are always designed to be as sparse as possible. This aids in the mathematical application of the axioms, there is less to prove if there are fewer axioms. To prove a function is a t-norm we need to show four things. Five would be harder. This sparseness, however means that the axioms do not always translate directly our intuitive notions. Thus, in most books, right after the axioms come proofs to show that the axioms imply the desired properties. For example let us prove that the axioms force the function t to generate the Table (6.1).

Theorem 3. $\mathsf{t}(0, 0) = \mathsf{t}(0, 1) = \mathsf{t}(1, 0) = 0$ *and* $\mathsf{t}(1, 1) = 1$.

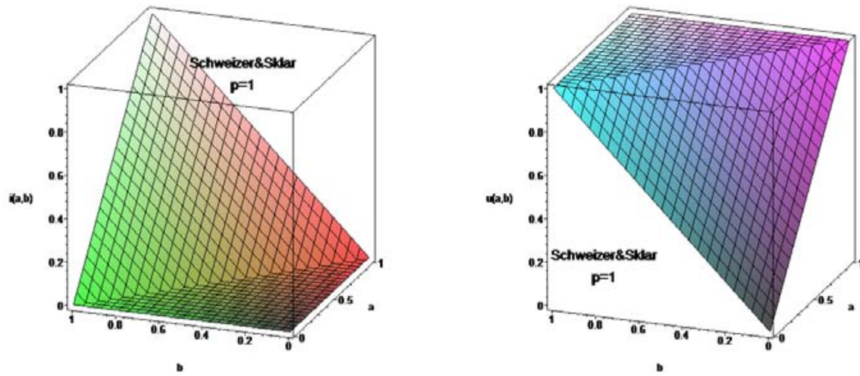


Figure 6.6.: Schweizer and Sklar t-norm and t-conorm functions for $p = 1$.

Proof. By axiom **(i1)** $\mathsf{t}(1, a) = a$ so set $a = 1$ and we have that $\mathsf{t}(1, 1) = 1$.

By axiom **(i1)** $\mathsf{t}(1, a) = a$ so set $a = 0$ and we have that $\mathsf{t}(1, 0) = 0$.

Commutativity **(i3)** says that if $\mathsf{t}(1, 0) = 0$ then $\mathsf{t}(0, 1) = 0$.

Finally $0 \leq 1$ and $1 \leq 1$ so axiom **(i2)** says that $\mathsf{t}(0, 0) \leq \mathsf{t}(1, 0) = 0$ but since $\mathsf{t}(a, b)$ cannot be negative the only value less than or equal to zero is zero we conclude that $\mathsf{t}(0, 0) = 0$. □

The largest t-norm is the minimum function and the smallest is the drastic product

$$\mathsf{t}_{\min}(a, b) = \begin{cases} a & \text{when } b = 1 \\ b & \text{when } a = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{6.8}$$

in the sense that if t is any t-norm then for any $a, b \in [0, 1]$

$$\mathsf{t}_{\min}(a, b) \leq \mathsf{t}(a, b) \leq \min(a, b). \tag{6.9}$$

One of the most commonly applied alternative t-norms is the algebraic product:

$$\mathsf{t}_{\mathsf{p}}(a, b) = a \cdot b, \tag{6.10}$$

which is calculated by simply multiplying a and b. Since multiplication is commutative and associative, axioms **i3** and **i4** are satisfied by t_{p}. Multiplication of nonnegative numbers is monotonic so axiom **i2** is satisfied. Finally $\mathsf{t}_{\mathsf{p}}(1, a) = 1 \cdot a = a$ so that axiom **i1** is satisfied. Thus t_{p} is a $\mathsf{t} - norm$.

Another common alternative to the standard intersection operator $\min$ is the bounded difference:

$$\mathsf{t}_{\mathsf{b}}(a, b) = \max(0, a + b - 1)\ . \tag{6.11}$$

Since $\mathsf{t}_{\mathsf{b}}(1, a) = \max(0, 1 + a - 1) = a$ the bounded difference satisfies axiom **i1**. The function $\max(0, a + b - 1)$ is obviously commutative (since $a + b - 1 = b + a - 1$) and since

$a+b-1$ and $\max$ are both monotone, their composition is monotone. Thus the bounded difference satisfies axioms **i3** and **i2**.

To show that the bounded difference is associative, and satisfies axiom **i4,** we will show that $\mathsf{t_b}(a,b,c) = \max(0, a+b+c-2)$. The following algebraic manipulation is based on the fact that, for any real number c, if $a < b$ then $a+c < b+c$ and if $a > b$ then $a+c > b+c$. This means that $\max(a,b)+c = \max(a+c, b+c)$.

First let us examine the left side of the equality in axiom **i4**:

$$\begin{aligned}
\mathsf{t_b}(a, \mathsf{t_b}(b,c)) &= \max(0, a + \mathsf{t_b}(b,c) - 1) &(6.12)\\
&= \max(0, a + \max(0, b+c-1) - 1) &(6.13)\\
&= \max(0, \max(a, a+b+c-1) - 1) &(6.14)\\
&= \max(0, \max(a-1, a+b+c-2)) &(6.15)\\
&= \max(0, a-1, a+b+c-2) &(6.16)
\end{aligned}$$

but, $a-1 \leq 0$ so that $\mathsf{t_b}(a, \mathsf{t_b}(b,c)) = \max(0, a+b+c-2)$.

Next we examine the right side of the equality in axiom **i4**:

$$\begin{aligned}
\mathsf{t_b}(\mathsf{t_b}(a,b),c) &= \max(0, \mathsf{t_b}(a,b) + c - 1) &(6.17)\\
&= \max(0, \max(0, a+b-1) + c - 1) &(6.18)\\
&= \max(0, \max(c, a+b+c-1) - 1) &(6.19)\\
&= \max(0, \max(c-1, a+b+c-2)) &(6.20)\\
&= \max(0, c-1, a+b+c-2) &(6.21)
\end{aligned}$$

but, $c-1 \leq 0$ so that $\mathsf{t_b}(\mathsf{t_b}(a,b),c) = \max(0, a+b+c-2)$.

We conclude that $\mathsf{t_b}(a, \mathsf{t_b}(b,c)) = \max(0, a+b+c-2) = \mathsf{t_b}(\mathsf{t_b}(a,b),c)$.

It can also be shown that the basic t-norms have the following order:

$$\mathsf{t}_{\min}(a,b) \leq \max\left(0, a+b-1\right) \leq ab \leq \min(a,b) \tag{6.22}$$

The most common t-norms are given in Table (6.2).

Name	t**-norm**
standard intersection	$\mathsf{t}(a,b) = \min(a,b)$
algebraic product	$\mathsf{t_p}(a,b) = ab$
bounded difference	$\mathsf{t_b}(a,b) = \max(0, a+b-1)$
drastic product	$\mathsf{t}_{\min}(a,b) = \begin{cases} a & \text{when } b=1 \\ b & \text{when } a=1 \\ 0 & \text{otherwise} \end{cases}$

Table 6.2.: Common t-norms.

It is sometimes convenient to think of a t-norm as binary operator, and use an in-line notation when applying it to a specific problem Thus instead of writing $\mathsf{t}(a,b)$ we would write $a \wedge_{\mathsf{t}} b$, or $a\ \mathsf{t}\ b$, or just $a \wedge b$, with the understanding that $\wedge$ is some intersection operator modeled by a t-norm t.

6.5. Fuzzy Set Unions

Similar ideas from the section on fuzzy set intersections lead to the following axiomatic skeleton for a function s to model the union operator. The union of two fuzzy sets must be a function that maps pairs of numbers in the unit interval into the unit interval, s: $[0,1] \times [0,1] \to [0,1]$. As is well known, functions known as *triangular conorms* or t-conorms, possess all the properties that are intuitively associated with fuzzy unions. They are characterized for all $a, b, c, d \in [0,1]$ by the following axioms:

(u1) Boundary condition — $\mathsf{s}(0, a) = a$.

(u2) Monotonicity — $\mathsf{s}(a, b) \leq \mathsf{s}(c, d)$ whenever $a \leq c$ and $b \leq d$.

(u3) Commutativity — $\mathsf{s}(a, b) = \mathsf{s}(b, a)$.

(u4) Associativity — $\mathsf{s}(a, \mathsf{s}(b, c)) = \mathsf{s}(\mathsf{s}(a, b), c)$.

The smallest t-conorm is the maximum function and the largest is the drastic sum sometimes called the drastic union:

$$\mathsf{s}_{\max}(a, b) = \begin{cases} a & \text{when } b = 0 \\ b & \text{when } a = 0 \\ 1 & \text{otherwise} \end{cases}, \tag{6.23}$$

in the sense that if s is any t-conorm then for any $a, b \in [0,1]$

$$\max(a, b) \leq \mathsf{s}(a, b) \leq \mathsf{s}_{\max}(a, b). \tag{6.24}$$

One of the most commonly applied alternative s-norms is the algebraic sum:

$$\mathsf{s}_{\mathsf{p}}(a, b) = a + b - ab, \tag{6.25}$$

which is also called the probabilistic sum.

Another common alternative to the standard union operator $\max$ is the bounded sum:

$$\mathsf{s}_{\mathsf{b}}(a, b) = \min(1, a + b)\ . \tag{6.26}$$

It can also be shown that the basic t-conorms have the following order:

$$\mathsf{s}_{\max}(a, b) \geq \min(1, a + b) \geq a + b - ab \geq \max(a, b) \tag{6.27}$$

The most common t-norms are given in Table (6.3).

It is sometimes convenient to think of a t-conorm (sometimes called an s-norm) as binary operator, and use an in-line notation when applying it to a specific problem Thus instead of writing $\mathsf{s}(a, b)$ we would write $a \vee_{\mathsf{s}} b$, or $a\ \mathsf{s}\ b$, or just $a \vee b$, with the understanding that $\vee$ is some union operator modeled by a t-conorm s.

Name	t-**conorm**
standard union	$\mathsf{s}(a,b) = \max(a,b)$
algebraic sum	$\mathsf{s_p}(a,b) = a + b - ab$
bounded sum	$\mathsf{s_b}(a,b) = \min(1, a+b)$
drastic sum	$\mathsf{s}_{\max}(a,b) = \begin{cases} a & \text{when } b = 0 \\ b & \text{when } a = 0 \\ 1 & \text{otherwise} \end{cases}$

Table 6.3.: Common t-conorms.

6.6. Residuum operator Omega operators

Let t be a continuous t-norm. Define the residuum operator, also called the ω (omega) operator generated by t, ω_{t}, for every $a, b \in [0,1]$ by the following definition

$$\omega_{\mathsf{t}}(a,b) = \sup\left\{x \in [0,1] \mid \mathsf{t}(a,x) \leq b\right\} \ . \tag{6.28}$$

The residuum operator operator is in one sense a model of material implication. Basically, we ask how much evidence (x) can we add to a before we break the threshold of belief b. It will play an important role in both the chapters on fuzzy relations, Ch. (8), and on fuzzy implication, Ch. (15).

6.7. Combinations of Operations

In classical set theory, the operations of intersection and union are dual with respect to the complement in that they satisfy the De Morgan laws.

1. The complement of the intersection of A and B equals the union of the complement of A and the complement of B.

2. The complement of the union of A and B equals the intersection of the complement of A and the complement of B.

$$(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}}$$
$$(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}}$$

Obviously, only certain combinations of t-norms, t-conorms, and fuzzy complements satisfy the duality. We say that a t-norm t and a t-conorm s are "dual with respect to a fuzzy complement c" if and only if

$$\mathsf{c}(\mathsf{t}(a,b)) = \mathsf{s}(\mathsf{c}(a), \mathsf{c}(b)) \tag{6.29}$$

and

$$\mathsf{c}(\mathsf{s}(a,b)) = \mathsf{t}(\mathsf{c}(a), \mathsf{c}(b)) \tag{6.30}$$

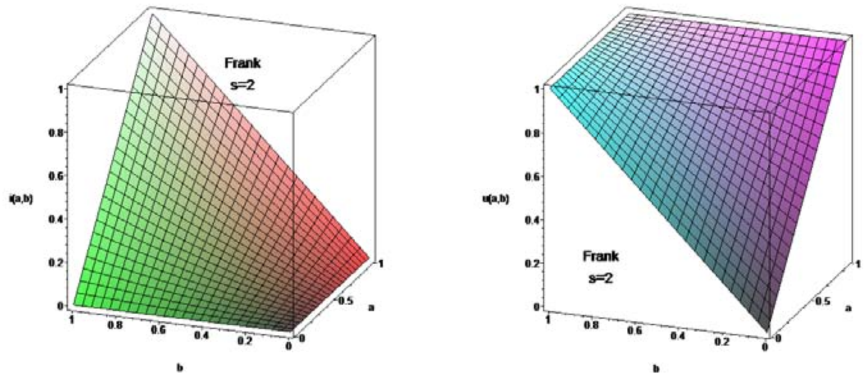


Figure 6.7.: Frank t-norm and t-conorm functions for $p = 2$.

These equations define the De Morgan laws for fuzzy sets. Let the triple $(\mathsf{t}, s, \mathsf{c})$ denote that t and s are dual with respect to c, and let any such triple be called a "dual triple".

The following t-norms and t-conorms are dual with respect to the Standard Fuzzy Complement c (i.e., dual triples):

t-norms	t-conorms	complement
$\min(a, b)$	$\max(a, b)$	c
ab	$a + b - ab$	c
$\max(0, a + b - 1)$	$\min(1, a + b)$	c
$I_{\min}(a, b)$	$U_{\max}(a, b)$	c

Table 6.4.: Triples of operators

Theorem 4. *The triples* $(\min, \max, \mathsf{c})$ *and* $(I_{\min}, U_{\max}, \mathsf{c})$ *are dual with respect to any fuzzy complement* c*.*

Theorem 5. *Given a* t*-norm* t *and an involutive fuzzy complement* c*, the binary operation* s *on* $[0, 1]$ *defined by*

$$\mathsf{s}(a, b) = \mathsf{c}(\mathsf{t}(\mathsf{c}(a), \mathsf{c}(b))) \tag{6.31}$$

for all a, b *in* $[0, 1]$ *is a* t*-conorm such that* $(\mathsf{t}, s, \mathsf{c})$ *is a dual triple.*

Theorem 6. *Given a* t*-conorm* s *and an involutive fuzzy complement* c*, the binary operation* t *on* $[0, 1]$ *defined by* $\mathsf{t}(a, b) = \mathsf{c}(s(\mathsf{c}(a), \mathsf{c}(b)))$ *for all* a, b *in* $[0, 1]$ *is a* t*-norm such that* $(\mathsf{t}, s, \mathsf{c})$ *is a dual triple.*

Theorem 7. *Given an involutive fuzzy complement* c *and an increasing generator* g *of* c*, the* t*-norm and* t*-conorm generated by* g *are dual with respect to* c*.*

Theorem 8. *Let* $(\mathsf{t}, s, \mathsf{c})$ *be a dual triple generated by Theorem 7. Then, the fuzzy operations* t*,* s*,* c *satisfy the law of excluded middle*

$$\mathsf{s}(a, \mathsf{c}(a)) = 1 \tag{6.32}$$

and the law of contradiction

$$\mathsf{t}(a, \mathsf{c}(a)) = 0. \tag{6.33}$$

Theorem. *Let (i,u,c) be a dual triple that satisfies the law of excluded middle and the law of contradiction. Then (i,u,c) does not satisfy the distributive laws. This means that* $\mathsf{t}(a, s(b, d))$ *is not equal to* $s(\mathsf{t}(a, b), \mathsf{t}(a, d))$ *for all* a, b, d *in* $[0, 1]$*.*

For detailed proofs of all these results see Klir and Yuan (1996).

We now can see the point of the rather tedious section on generator functions. If we would like a dual triple, which makes the algebraic manipulation of fuzzy sets a little simpler, then we can create such a dual triple from one generating function g. In application, the complement is the simplest function to analyze. Suppose we can generate some data that gives us a feel for how people feel subjectively about membership and non-membership of some object in a fuzzy set. If we can fit a Sugeno or Yager complement to this data, we can retrieve its generating function g and then use it to generate corresponding (dual triple) union, s, and intersection, t, operators. The homework at the end of this chapter describes how such an experiment could be performed.

6.8. Averaging Operator

The averaging operators are a third class of binary operators used to average its arguments a and b. Since all intersection operators produce results that are below the minimum of a and b and all union operators produce results that are greater than the maximum of a and b, there is a large range of values that are excluded by these two classes of operators. Into this gap we now introduce averaging operators $\mathsf{h}(a, b)$. These operators do not correspond exactly to any logical connective, the way intersection operators model and and union operators model or. These averaging operators take two arguments and produce a result that greater than or equal to the $\min(a, b)$ and less than or equal to $\max(a, b)$. An averaging operator is a function $\mathsf{h} : [0, 1] \times [0, 1] \to [0, 1]$ such that following axioms hold.

(h1) Idempotency — $\mathsf{h}(a, a) = a$.

(h2) Monotonicity — $\mathsf{h}(a, b) \leq \mathsf{h}(c, d)$ whenever $a \leq c$ and $b \leq d$.

(h3) Commutativity — $\mathsf{h}(a, b) = \mathsf{h}(b, a)$.

(h4) Continuity — h is a continuos function.

Name	Operator	Generator
arithmetic mean	$\frac{a+b}{2}$	x
generalized p-mean	$\sqrt[p]{\frac{a^p+b^p}{2}}$	x^p
harmonic mean	$\frac{2ab}{a+b}$	$\frac{1}{x}$
geometric mean	$\sqrt{xy}$	$\log x$
dual of geometric mean	$1-\sqrt{(1-x)(1-y)}$	$\log(1-x)$

Table 6.5.: Averaging operators.

The following properties could have been listed as a condition but since they are consequences of the previous axioms it could also be stated as a lemma. This is another example of keeping an axiomatic skeleton sparse. If we added the following two conditions as axioms student would have to show six things in a homework problem to show that an operator was an averaging operator. Instead they only have to show four things, and that is certainly easier.

(g5) Extremes — $\mathsf{h}(0,0)=0$ and $\mathsf{h}(1,1)=1$.

(g6) Boundary conditons — $\min(a,b)\leq \mathsf{h}(a,b)\leq \max(a,b)$.

Averaging operators allow for an interaction between the values of two fuzzy sets. It allows the resultant averaged value to be better than the worst case but less than the best case. In fact the average value is often right in the middle, which should come as no surprise. However there are other averaging operators beside the geometric mean, such as the harmonic mean. Assume f is any continuous strictly monotone function (this means that it is always increasing or always decreasing). It is a fact that all continuous strictly monotone functions have inverses so we know that f^{-1} exists. Then

$$\mathsf{h}(a,b)=f^{-1}\left[\frac{f(a)+f(b)}{2}\right]$$

is called a quasi-arithmetic means. Such a function h is always an averaging operator Aczél (1966). Let $\alpha\in[0,1]$ then

$$\mathsf{h}(a,b)=f^{-1}\left[\alpha f(a)+(1-\alpha)f(b)\right]$$

is a more general form of quasi-arithmetic operator Aczél (1966). Some of the more important averaging operators are given in Table (6.5).

6.9. Aggregation Operations

The idea of an averaging operator can be extended to m-ary aggregation operators. Since averaging operations are in general not associative, they must be defined as functions of m arguments ($m\geq 2$). That is, an averaging operation h is a function of the form

$$\mathsf{h} : [0,1]^m \to [0,1]. \tag{6.34}$$

Averaging operations are characterized by the following set of axioms:

(h1) Idempotency — for all $a \in [0,1]$,

$$\mathsf{h}(a, a, a, \ldots, a) = a. \tag{6.35}$$

(h2) Monotonicity — for any pair of m–tuples of real numbers in $[0,1]$, $\langle a_1, a_2, a_3, \ldots, a_m \rangle$ and $\langle b_1, b_2, b_3, \ldots, b_m \rangle$, if $a_k \le b_k$ for all $k \in \mathbb{N}_m$ then

$$\mathsf{h}(a_1, a_2, a_3, \ldots, a_m) \le \mathsf{h}(b_1, b_2, b_3, \ldots, b_m). \tag{6.36}$$

It is significant that any function h that satisfies these axioms produces numbers that, for any m–tuple $\langle a_1, a_2, a_3, \ldots, a_m \rangle \in [0,1]^m$, lie in the closed interval defined by the inequalities

$$\min(a_1, a_2, a_3, \ldots, a_m) \le \mathsf{h}(a_1, a_2, a_3, \ldots, a_m) \le \max(a_1, a_2, a_3, \ldots, a_m). \tag{6.37}$$

The $\min$ and $\max$ operations qualify, due to their idempotency, not only as fuzzy counterparts of classical set intersection and union, respectively, but also as extreme averaging operations.

An example of a class of symmetric averaging operations are *generalized means*, which are defined for all m–tuples $\langle a_1, a_2, a_3, \ldots, a_m \rangle$ in $[0,1]^m$ by the formula

$$\mathsf{h}_p(a_1, a_2, a_3, \ldots, a_m) = \frac{1}{m}(a_1^p + a_2^p + a_3^p + \cdots + a_m^p)^{\frac{1}{p}}, \tag{6.38}$$

where p is a parameter whose range is the set of all real numbers excluding 0. For $p = 0$, h_p is not defined; however for $p \to 0$, h_p converges to the well known geometric mean. That is, we take

$$\mathsf{h}_0(a_1, a_2, a_3, \ldots, a_m) = (a_1 a_2 a_3 \ldots a_m)^{\frac{1}{m}}. \tag{6.39}$$

For $p \to -\infty$ and $p \to \infty$, h_p converges to the $\min$ and $\max$ operations, respectively.

Assume again that f is any continuous strictly monotone function, then

$$\mathsf{h}(a_1, a_2, .., a_m) = f^{-1}\left[\frac{1}{m}\sum_{i=1}^{m} f(a_i)\right]$$

is still called a quasi-arithmetic means . Let $\langle w_1, w_2, ..., w_m \rangle$ be weights with $w_i \in [0,1]$ then

$$\mathsf{h}(a_1, a_2, .., a_m) = f^{-1}\left[\sum_{i=1}^{m} w_i f(a_i)\right]$$

is a more general form of quasi-arithmetic operator Aczél (1966).

6.9.1. OWA operators

Yager introduced ordered weighted averaging(OWA) operators in Yager (1988). They are by nature averaging operators that treat a fuzzy set in its possibility theory interpretation. Let $\mathbf{a} = \langle a_1, a_2, a_3, \ldots, a_m \rangle$ be an m-dimensional vector of values and let $\mathbf{w} = \langle w_1, w_2, w_3, \ldots, w_m \rangle$ be an m-dimensional vector of weights, with both $a_i \in [0,1]$ and $w_i \in [0,1]$, for $1 \leq i \leq m$. Define the vector $\mathbf{b} = \langle b_1, b_2, b_3, \ldots, b_m \rangle$ to be the vector $\mathbf{a}$ sorted in decreasing order of magnitude, so that $b_i \geq b_{i+1}$ then the OWA average of $\mathbf{a}$ is

$$OWA_{\mathbf{w}}(\mathbf{a}) = \sum_{i=1}^{m} w_i b_i.$$

At first it would seam that OWA operators are very artificial. However. let us examine three special cases of OWA operators

Let $w^* = \langle 1, 0, 0, ..., 0 \rangle$ then

$$OWA_{w^*}(\mathbf{a}) = b_1 = \max\left[a_1, a_2, a_3, \ldots, a_m\right] \ .$$

Let $w_* = \langle 0, 0, 0, ..., 1 \rangle$ then

$$OWA_{w_*}(\mathbf{a}) = b_m = \min\left[a_1, a_2, a_3, \ldots, a_m\right] \ .$$

Let $w_* = \left\langle \frac{1}{m}, \frac{1}{m}, \frac{1}{m}, ..., \frac{1}{m} \right\rangle$ then

$$OWA_{w_*}(\mathbf{a}) = \frac{1}{m} \sum_{i=1}^{m} b_i = \frac{1}{m} \sum_{i=1}^{m} a_i \ .$$

Thus OWA operators allow us to perform a delicate mix of values emphasizing either large values in $\mathbf{a}$ by making w_i big for low values of i and tiny for higher values of i or vice versa.

6.10. Notes

Most of the important early work on t-norms comes form early, and quite complex, works on probabilistic metric spaces. The term is originalay due to Menger (1942). Most of the developmental work was done in Schweizer and Sklar (1961) and Schweizer and Sklar (1963). The book Schweizer and Sklar (1983) provides a complete development of the subject. Characterization of the union and intersection operators for fuzzy set theory was presented in Bellman and Giertz (1973). An excellent overview of aggregation operators is in Dubois and Prade (1985).

6.11. Homework

Let us define the universal sets

$$X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}, \tag{6.40}$$

$$Y = \{a, b, c\}, \tag{6.41}$$

and

$$Z = \{\alpha, \beta, \gamma, \delta\}\,. \tag{6.42}$$

Given the fuzzy sets

$$A(x) = \frac{x}{10}$$
$$B(x) = \frac{|x-5|}{5}$$
$$C(x) = \begin{cases} \frac{9-|2x-9|}{8} & 1 \le x \le 8 \\ 0 & otherwise \end{cases}$$
$$D(x) = 0.8$$
$$E(x) = \{\langle 1, 0.2\rangle, \langle 2, 0.6\rangle, \langle 3, 0.4\rangle\}$$
$$F(y) = \{\langle a, 0.3\rangle, \langle b, 0.7\rangle, \langle c, 0.9\rangle\}$$

and the definition of two different fuzzy relations.

R	a	b	c
1	1.0	0.5	0.1
2	0.4	1.0	0.2
3	0.6	0.5	1.0

S	α	β	δ
a	0.9	0.5	0.1
b	0.4	0.7	0.2
c	0.6	0.5	0.9

(6.43)

try to answer the following questions.

1. What is the Sugeno complement of A if $\lambda = -1$?
2. What is the Sugeno complement of A if $\lambda = 2$?
3. What is the Sugeno complement of A if $\lambda = 10$?
4. What is the Yager complement of A if $\omega = 0.5$?
5. What is the Yager complement of A if $\omega = 2$?
6. What is the threshhold complement of A if $t = 0.0$?
7. What is the threshhold complement of A if $t = 0.5$?
8. What is the threshhold complement of A if $t = 1.0$?
9. What is the Sugeno complement of B if $\lambda = -1$?
10. What is the Sugeno complement of B if $\lambda = 2$?
11. What is the Sugeno complement of B if $\lambda = 10$?
12. What is the Yager complement of B if $\omega = 0.5$?
13. What is the Yager complement of B if $\omega = 2$?

14. What is the threshhold complement of B if $t = 0.0$?
15. What is the threshhold complement of B if $t = 0.5$?
16. What is the threshhold complement of B if $t = 1.0$?
17. What is a t-norm? Explain this in English.
18. What are the extreme t-norms? Explain why they are considered extreme.
19. What is a t-conorm? Explain this in English.
20. What are the extreme t-conorms? Explain why they are considered extreme.
21. What is the $A \cap B$ using the t-norm drastic product.
22. What is the $A \cap B$ using the t-norm bounded difference.
23. What is the $A \cap B$ using the t-norm algebraic product.
24. Contrast the previous three results.
25. What is the $A \cap C$ using the t-norm drastic product.
26. What is the $A \cap C$ using the t-norm bounded difference.
27. What is the $A \cap C$ using the t-norm algebraic product.
28. Contrast the previous three results.
29. What is the $B \cap C$ using the t-norm drastic product.
30. What is the $B \cap C$ using the t-norm bounded difference.
31. What is the $B \cap C$ using the t-norm algebraic product.
32. Contrast the previous three results.
33. What is the $A \cup B$ using the t-conorm drastic sum.
34. What is the $A \cup B$ using the t-conorm bounded sum.
35. What is the $A \cup B$ using the t-conorm algebraic sum.
36. Contrast the previous three results.
37. What is the $A \cup C$ using the t-conorm drastic sum.
38. What is the $A \cup C$ using the t-conorm bounded sum.
39. What is the $A \cup C$ using the t-conorm algebraic sum.
40. Contrast the previous three results.
41. What is the $B \cup C$ using the t-conorm drastic sum.
42. What is the $B \cup C$ using the t-conorm bounded sum.

43. What is the $B \cup C$ using the t-conorm algebraic sum.

44. Contrast the previous three results.

45. Is $g(x) = x^2$ a generating function? If so what complement, t-norm and t-conorm are generated by $g(x)$?

46. Is $g(x) = \log_2(1 + x)$ a generating function? If so what complement, t-norm and t-conorm are generated by $g(x)$?

47. Is $g(x) = e^x - 1$ a generating function? If so what complement, t-norm and t-conorm are generated by $g(x)$?

7. Fuzzy Numbers

The real building block of applied fuzzy set theory is the fuzzy number.

Fig. (7.1) shows the evolution from the simple natural number, 135, with a representational graph $\mathsf{A}(x)$ that is one at $x = 135$ and zero elsewhere, to an interval containing 135 with representational graph $\mathsf{B}(x)$ which is one on the interval $133 \leq x \leq 137$, to a fuzzy number about 135 with representational graph $\mathsf{C}(x)$ which rises in a straight line from the point $\langle 131, 0\rangle$ to the point $\langle 135, 1\rangle$ and then drops back down to the point $\langle 139, 0\rangle$, and finally the fuzzy interval near 135 with representational graph $\mathsf{D}(x)$ that rises in a straight line from the point $\langle 131, 0\rangle$ to the point $\langle 133, 1\rangle$ stays level at the height of 1.0 to the point $\langle 137, 1\rangle$ and then drops down to the point $\langle 139, 0\rangle$. Recently the distinction between a fuzzy number and a fuzzy interval has been dropped and both have been referred to as fuzzy numbers, but the terminology fuzzy interval still occurs intermittently in the literature.

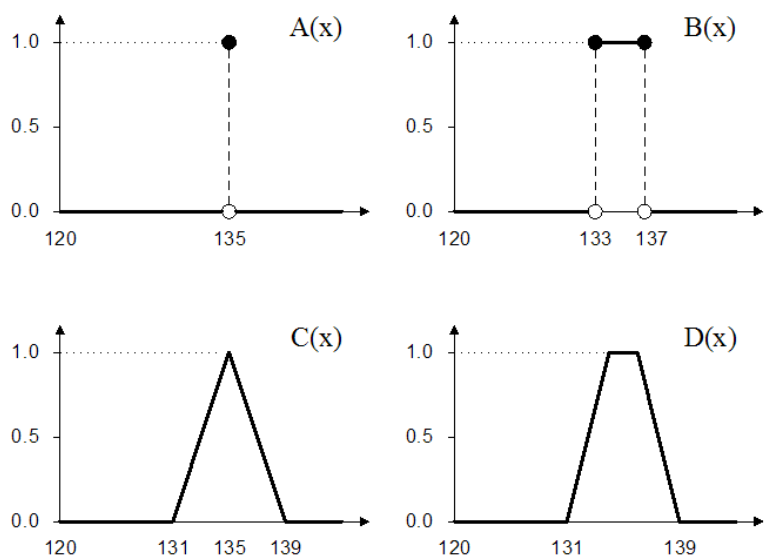


Figure 7.1.: From the real numbers to the fuzzy intervals: (a) real number, (b) crisp interval, (c) fuzzy number, and (d) fuzzy interval.

7.1. Fuzzy Numbers

A fuzzy number, T, is a bounded, convex, and normal fuzzy set defined upon the reals.

Before we explore the technical details let us start with a simple example. Eq. (7.1) and (7.2) define two fuzzy numbers A and B. The fuzzy number A represents the vague concept about one and B represents the vague concept about two. A fuzzy set is determined by its membership function so that the equation for fuzzy set A is

$$\mathsf{A}(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 < x \leq 2 \end{cases} \tag{7.1}$$

and the equation for fuzzy set A is

$$\mathsf{B}(x) = \begin{cases} x-1 & 1 \leq x \leq 2 \\ 3-x & 2 < x \leq 3 \end{cases} \tag{7.2}$$

If we examine a graph of A and A (Figs. (7.2))we see that the fuzzy numbers are represented by triangularly shaped functions made up of two line segments. Note also, that we suppressed mentioning the additional restriction that $\mathsf{A}(x) = 0$ if x is outside the closed interval $[0, 2]$ and that $\mathsf{B}(x) = 0$ if x is outside the closed interval $[1, 3]$. From this moment on, we shall assume that every fuzzy number has zero as its membership grade for every domain value of x in the universe X not explicitly listed as having a nonzero membership grade. The fuzzy number A expresses the notion that about one includes any domain value greater than zero and less than two, but also indicates that we are more confident that 1.1 is about one (0.9 confident) than we are that .2 is about one (0.2 confident) and that 2.1 is not about one at all (0.0 confidence). This is somewhat related to notions like accuracy and precision but on a sliding scale. For instance a ruler that is marked in centimeters must be read to the nearest centimeter. A measurement of 34 centimeters says that the exact value is in the interval [33.5,34.5), however all values in this range are equally 34 centimeters.

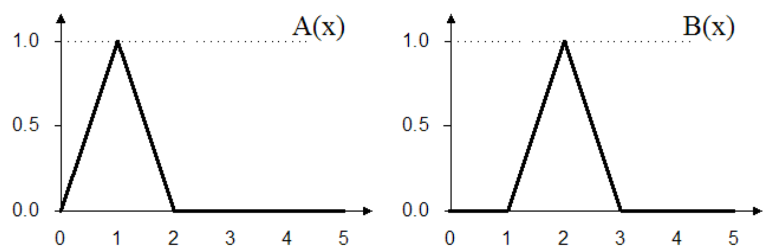


Figure 7.2.: The fuzzy numbers $A(x)$, about one and $B(x)$, about two.

Fuzzy numbers are one of the focal concepts of this book. How we design, manipulate, measure, utilize and compare fuzzy numbers are the subjects that concern most the remaining chapters on fuzzy set theory.

For a start on using fuzzy numbers, consider the problem of adding two fuzzy numbers, say A and B. The result should be another fuzzy number and if the fuzzy numbers A and B properly represent the concepts about one and about two then their sum should represent about three. In this example of the addition of two fuzzy numbers A and B we are already being exposed to the true power of fuzzy numbers. Nothing before fuzzy set theory was designed to handle the everyday quantities of human existence like around a dozen. When the recipe calls for two large onions and all we have is a bag of small onions we are not daunted. If we are hungry we use three, maybe four small onions as an approximation. (Or we go get takeout.). Chapters (16) and (9)

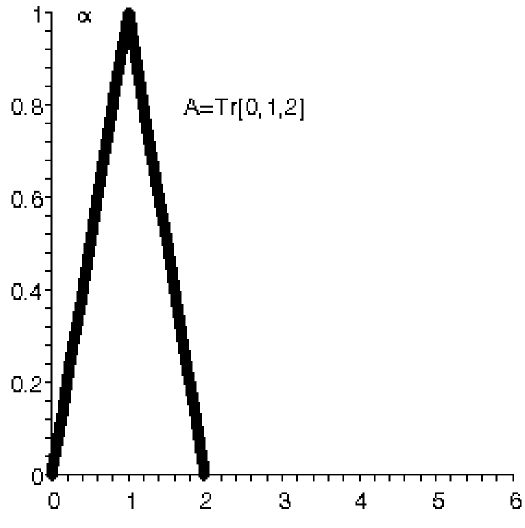


Figure 7.3.: The fuzzy triangular number $A = Tr[0, 1, 2]$.

will focus on how we get and use about one, very old, and other vague concepts. Right now, we will present some of the technical details of fuzzy numbers before resuming our discussion on how to add two fuzzy numbers.

Here is the technical definition of a fuzzy number.

Definition 20 (fuzzy number)**.** A fuzzy number is a fuzzy set with domain $\mathbb{R}$, the real numbers, that is

1. normal, some element has membership grade one,

2. bounded, the support of the fuzzy set is a bounded interval, and

3. convex, essentially, every α–cut is a closed interval for positive α.

For the mathematically inclined, a fuzzy set T is termed normal if there exists an $x \in X$ such that $\mathsf{T}(x) = 1$. A fuzzy set T is convex if every α–cut T^{α}, for $\alpha \in (0, 1]$, is a convex subset of the domain, i.e., a continuous interval. An equivalent definition of convexity requires that for any λ, $0 \leq \lambda \leq 1$, that $\min[\mathsf{T}(x), \mathsf{T}(y)) \leq \mathsf{T}(\lambda x + (1 - \lambda)y]$. Finally the support, or strong α–cut at zero must be bounded, that is $\mathsf{T}^{0+} = [a, b]$ with $a \leq b$ and neither a nor b is permitted to be infinite.

Remark 4. Many earlier works had fuzzy numbers being unimodal. That meant that the x such that $\mathsf{T}(x) = 1$ had to be unique. These works distinguished between fuzzy numbers (unimodal) and fuzzy intervals where $\mathsf{T}(x) = 1$ over an interval $[l, r]$. However, the similarities between the two entities were so great that the distinction was soon abandoned.

7.1.0.1. Triangular fuzzy number

A triangular fuzzy number, Tr, is named for its shape. Its membership function is given by two line segments, one rising from the point $\langle a, 0\rangle$ to $\langle m, 1\rangle$ and the second falling from $\langle m, 1\rangle$ to $\langle b, 0\rangle$. Its domain is the closed interval $[a, b]$. A triangular fuzzy number can be specified by the ordered triple $\langle a, m, b\rangle$ with $a \leq m \leq b$ and its membership function is:

$$\mathsf{Tr}\,[a, m, b]\,(x) = \begin{cases} \frac{x-a}{m-a} & a \leq x \leq m \\ \frac{x-b}{m-b} & m < x \leq b \end{cases}$$

Remark 5. When an ordered pair $\langle x, y\rangle$ is used as an argument to a function of two variables, it is almost always written as $f(x, y)$ and not as $f(\langle x, y\rangle)$. The extra brackets only make the expression more difficult to interpret. Similarly, parameters are usually enclosed in square brackets, $[p]$, and $\mathsf{Tr}\,[\langle a, m, b\rangle]$ is abbreviated to $\mathsf{Tr}\,[a, m, b]$. However the parameters of the triangular fuzzy number, and the other fuzzy numbers to come, are in order, and the square brackets do not represent a closed interval.

7.1.0.2. Trapezoidal fuzzy number

A trapezoidal fuzzy number, Tp, can be specified by an ordered quadruple $\langle a, l, r, b\rangle$ with $a \leq l \leq r \leq b$ and a membership function consisting of three line segments. The first segment rises from $\langle a, 0\rangle$ to $\langle l, 1\rangle$, the second segment is a horizontal line that has a constant value of one and that stretches from $\langle l, 1\rangle$ to $\langle r, 1\rangle$, and the third segment drops from $\langle r, 1\rangle$ to $\langle b, 0\rangle$. The membership functions for a trapezoidal fuzzy numbers is:

$$\mathsf{Tp}\,[a, l, r, b]\,(x) = \begin{cases} \frac{x-a}{l-a} & a \leq x \leq l \\ 1 & l < x < r \\ \frac{x-b}{r-b} & r \leq x \leq b \end{cases}$$

Other standard fuzzy number are the differentiable piecewise quadratic and Gaussian bell shaped numbers.

7.1.0.3. Differentiable piecewise quadratic

A differentiable piecewise quadratic numbers Tq consist of four quadratic pieces and are parameterized by five values a, l, m, r, and b. The values a and b are the left and right hand limits of the support, m is the mean value (core) where Tq assumes the value one, and l and r are the left and right hand crossover points (points of inflection). If $\mathsf{Tq}\,\langle a, l, m, r, b\rangle$ is a *continuous piecewise quadratic number* then the membership function of T_q is given by:

$$\mathsf{Tq}\,[a, l, m, r, b]\,(x) = \begin{cases} \frac{1}{2}\left(\frac{a-x}{a-l}\right)^2 & a \leq x \leq l \\ 1 - \frac{1}{2}\left(\frac{m-x}{m-l}\right)^2 & l < x \leq m \\ 1 - \frac{1}{2}\left(\frac{m-x}{m-r}\right)^2 & m < x \leq r \\ \frac{1}{2}\left(\frac{b-x}{b-r}\right)^2 & r < x \leq b \end{cases}$$

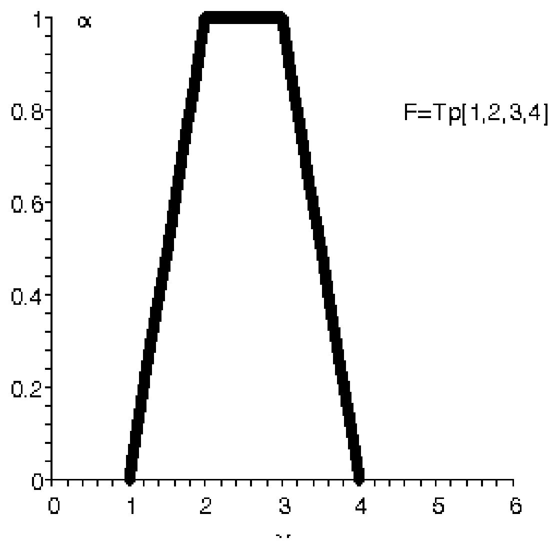


Figure 7.4.: The fuzzy tapezoidal number $F = Tp[1, 2, 3, 4]$.

7.1.0.4. Gaussian numbers

Gaussian numbers are parameterized by m the mean, s the spread and γ the scale parameter that adjusts the height. The membership function of a *bell shaped* fuzzy number is:

$$\mathsf{Tb}\left[m, s, \gamma\right](x) = \gamma\, e^{-(x-m)^2/s^2}$$

7.1.0.5. $L - R$ fuzzy number

An $L - R$ fuzzy number Dubois and Prade ((1980a,)) is a unimodal fuzzy number on the reals that can be described in terms of two reference function, the left hand reference function, ${}^{L}F$, and the right hand function, ${}^{R}F$. A unimodal fuzzy set has its maximum value at a unique value m of the domain X. This insures that if $T(x) = 1$ then $x = m$. A reference function ${}^{L}F$ or ${}^{R}F$ is a function that is monotone non- decreasing on the interval $(-\infty, 0)$ and monotone non-increasing on the interval $(0, \infty)$ and such that ${}^{L}F(0) = {}^{R}F(0) = 1$. With these conventions an $L-R$ fuzzy number can be described in terms of an ordered triple

$$\mathsf{Tlr} = \left\langle \langle m, u, v\rangle, {}^{L}F, {}^{R}F \right\rangle$$

where ${}^{L}F$ and ${}^{R}F$ are the reference functions, m is the mean of the fuzzy number and u and v are the left and right–hand spread of the function. The formal definition of

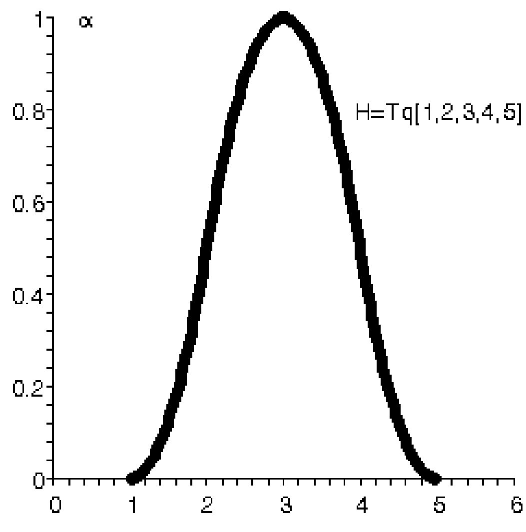


Figure 7.5.: The piecewise quadratic fuzzy number $H = Tq\,[1, 2, 3, 4, 5]$.

the $L - R$ fuzzy number is

$$\mathsf{Tlr}(x) = \begin{cases} {}^{R}F(\frac{x-m}{v}) & x \geq m \\ {}^{L}F(\frac{m-x}{u}) & x < m \end{cases}$$

Almost all fuzzy numbers can be expressed as $L - R$ fuzzy numbers with an appropriate choices for the left and right reference function ${}^{L}F$ and ${}^{R}F$ as well as for the parameters; the mode m, left spread u. and right spread v.

7.1.0.6. Impulse fuzzy number

Finally we mention the impulse fuzzy numbers also called a fuzzy point number. The impulse number $\mathsf{Ti}\,[m]$ has membership function

$$\mathsf{Ti}\,[m]\,(x) = \begin{cases} 1 & x = m \\ 0 & \text{otherwise} \end{cases} \quad . \tag{7.3}$$

An impulse number can be considered a degenerate case of a triangular (and other types) of fuzzy number since $\mathsf{Ti}\,[m] = \mathsf{Tr}\,\langle m, m, m\rangle$. This is also called a fuzzy singleton in the literature of fuzzy set theory.

7.2. S–shaped fuzzy sets

The bounded support requirement for a fuzzy number means that what goes up must come down. While this is an important requirement for fuzzy numbers, many monotone functions that asymptotically approach one are treated as if they are fuzzy num-

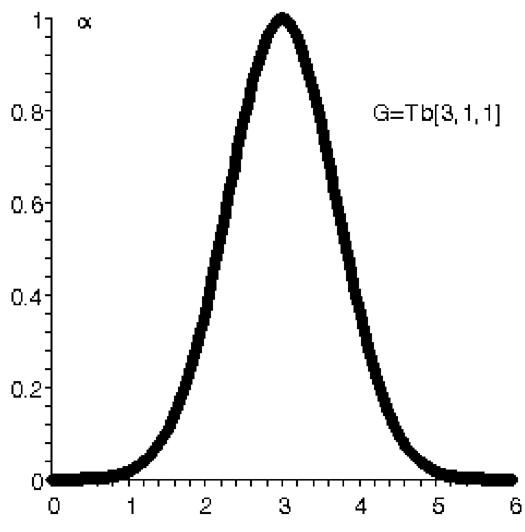


Figure 7.6.: The fuzzy bell number $G = Tb\,[3, 1, 1]$.

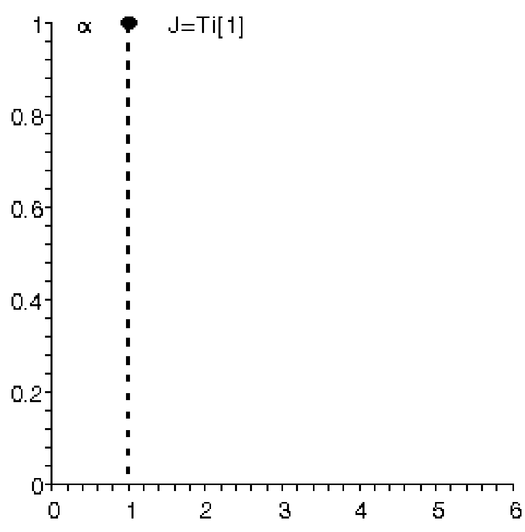


Figure 7.7.: The fuzzy singleton $Ti\,[1]$.

bers, though, technically, they are not. An asymptotic function gets very close to a limiting value (which is usually zero or one in this book) but never quite reaches it, as shown in Fig. (7.9).

Monotone increasing and monotone decreasing functions as illustrated in Fig. (7.8) are often called an s–shaped fuzzy number, which is an abuse of the language that occurs all too often. An increasing s–shaped fuzzy set S may consist of two quadratic pieces, line segments, exponential curves, or any monotonic function that achieves or is asymptotic to 1 at one end of its domain set, and achieves or is asymptotic to 0 at the other end of its domain set. Assume in all the following definitions of this section that $a \le l \le m \le r \le b$ are all real numbers.

Some examples of increasing s–shaped fuzzy sets are the quadratic s-shaped fuzzy sets Sqi:

$$\mathsf{Sqi}\,[a,l,m]\,(x) = \begin{cases} \frac{1}{2}\left(\frac{a-x}{a-l}\right)^2 & a \le x \le l \\ 1-\frac{1}{2}\left(\frac{m-x}{m-l}\right)^2 & l < x \le m \end{cases}$$

and the linear s-shaped sets Sli:

$$\mathsf{Sli}\,[a,l,r,b]\,(x) = \begin{cases} 0 & a \le x \le l \\ \frac{x-l}{r-l} & l < x \le r \\ 1 & r < x \le b \end{cases}$$

Some examples of decreasing s–shaped fuzzy numbers are the quadratic s-shaped fuzzy sets Sqd:

$$\mathsf{Sqd}\,[m,r,b]\,(x) = \begin{cases} 1-\frac{1}{2}\left(\frac{m-x}{m-r}\right)^2 & m \le x \le r \\ \frac{1}{2}\left(\frac{b-x}{b-r}\right)^2 & r < x \le b \end{cases}$$

and the linear s-shaped sets Sld:

$$\mathsf{Sld}\,[a,l,r,b]\,(x) = \begin{cases} 1 & a \le x \le l \\ \frac{x-r}{l-r} & l < x \le r \\ 0 & r < x \le b \end{cases}$$

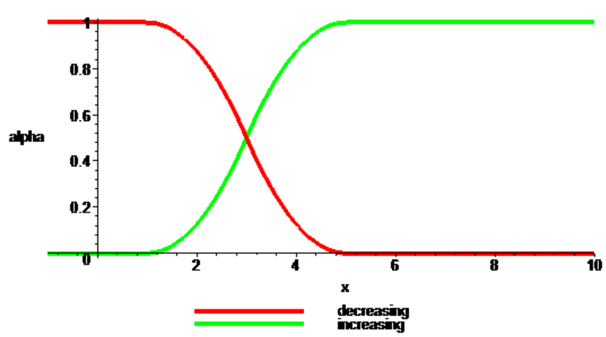


Figure 7.8.: An increasing and a decreasing s-shaped fuzzy set.

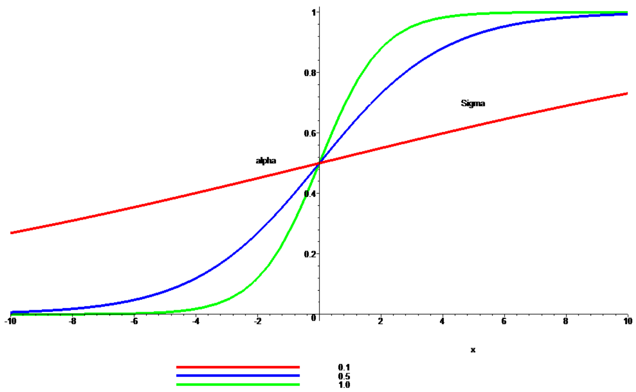


Figure 7.9.: Sigmoid fuzzy sets.

Other extremely popular increasing s-shaped fuzzy sets are those generated by the sigmoid functions,

$$\mathsf{S}\sigma\left[\beta\right](x) = \left(1 + e^{-\beta x}\right)^{-1} .$$

with $\beta \geq 0$. This curve is illustrated in Fig. (7.9), for various values of $\beta > 0$. If β is less than zero then the s-shaped fuzzy set becomes decreasing.

7.3. Fuzzy Arithmetic

7.3.1. The extension principle

The most important tool in all of fuzzy sets is the extension principle. It is important because it provides the connection between fuzzy sets and all the functions, operations, and tools of classical mathematics.

For example suppose we define a fuzzy number $\mathsf{A} = \mathsf{Tr}\left[0, 1, 2\right]$ and another fuzzy number $\mathsf{B} = \mathsf{Tr}\left[1, 2, 3\right]$.

It is a natural question to ask what is the sum of these two fuzzy numbers? Also what is their difference, product and quotient? Furthermore, how do we define functions upon fuzzy numbers? Fuzzy numbers would be of little use if there were no answers for these questions, or if there were not some general method to extend any function or operation of classical mathematics to the realm of fuzzy numbers. The extension principle provides this ability and more.

First it is somewhat obvious that the sum of two fuzzy numbers would again be a fuzzy number. A fuzzy number such as A models the concept about one and B models about two. If we add about one to about two it would be surprising if we got an exact value, in fact we would expect the answer to be about three. Another observation we can make is the following. Since we are unsure as to how oneish A is and how twoish B is, we are doubly unsure how threeish their sum C would be.

Thus the sum of A and B should be another fuzzy number C and we need a formula to define the membership grade of x in C based upon the definitions of A and B.

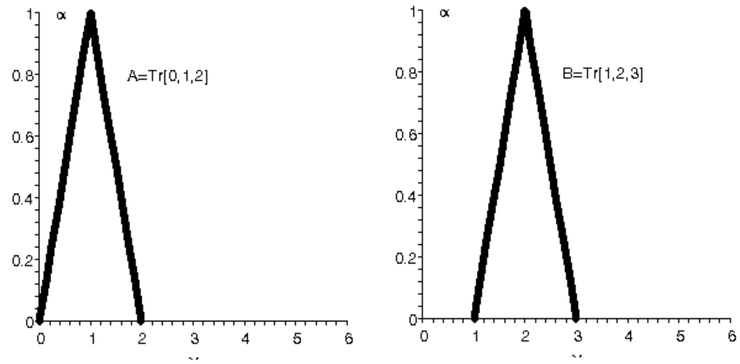


Figure 7.10.: The fuzzy numbers $A = Tr\,[0, 1, 2]$ and $B = Tr\,[1, 2, 3]$.

The proper definition is

$$\mathsf{C}(z) = \max_{x+y=z} \min[\mathsf{A}(x), \mathsf{B}(y)]. \tag{7.4}$$

which gives us the membership function of $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$. In this example we are assuming that x, y, and z are all ranging over the real numbers $\mathbb{R}$. We use the notation $\oplus$ for the addition of fuzzy numbers because it is important to remember that $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$ cannot be gotten by adding the two membership functions of A and B. It is more on the order of a convolution than a simple addition. The result turns out to be another triangular fuzzy number,

$$\mathsf{C}(x) = \begin{cases} \frac{x-1}{2} & 1 \leq x \leq 3 \\ \frac{5-x}{2} & 3 < x \leq 5 \end{cases} \tag{7.5}$$

The formulas (7.4) and (7.5) are produced by the extension principle. However the extension principle is a very general result that can be applied to many other situations besides the addition of two fuzzy numbers. It can be applied to any of the operators of ordinary arithmetic to provide formulas for the sum, difference, product and quotient of fuzzy numbers. It can be applied to any crisp numerical relation or function in mathematics to provide an analogous fuzzy numeric relation or function.

Let X be the Cartesian product of n sets, $\mathbf{X} = X_1 \times X_2 \times X_3 \cdots X_n$. Let A_1, A_2, A_3,$\cdots$,A_n be n fuzzy sets defined upon the universes X_1, X_2, X_3 $\cdots$ X_n respectively. Suppose further that f is a mapping from $\mathbf{X}$ to some set Y, $y = f(x_1, x_2, x_3, ..., x_n) = f(\mathbf{x})$. Lastly, by $f^{-1}(y)$ we denote that subset of $\mathbf{X}$ that is mapped by the function f to $y \in Y$, i.e., $f^{-1}(y) = \{x \mid f(x) = y\}$. We can now define a fuzzy set B with domain Y, induced by the mapping f and the fuzzy sets $A_1, A_2, A_3, \ldots, A_n$, via the extension principle

$$B(y) = \sup_{\mathbf{x} \in f^{-1}(y)} \min[A_1(x_1), A_2(x_2), \cdots, A_n(x_n)] \tag{7.6}$$

for all $y \in Y$. Note that f need not be a function, only a relation.

Thus the definition of the membership function of the sum of B and B, $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$, given above is due to the application of the extension principle upon the fuzzy numbers A and B and the binary arithmetic function of addition.

With the help of the extension principle we can now define a fuzzy arithmetic of

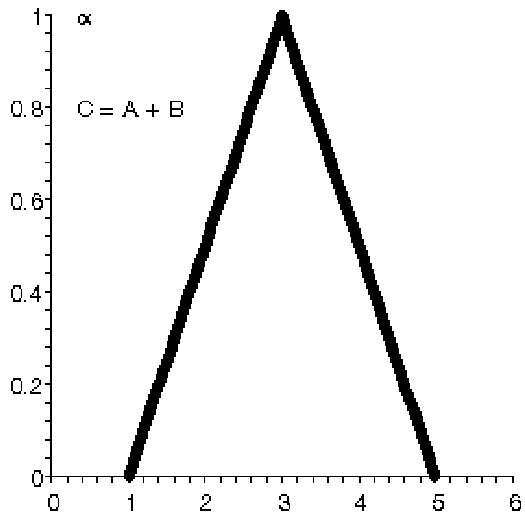


Figure 7.11.: The fuzzy set $C(x)$, about three, is the sum of $A(x)$, about one and $B(x)$, about two.

fuzzy numbers. Recalling the definitions of A and B the extension principle gave us $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$ with membership function $\mathsf{C} = Tr\,[1,3,5]$:

$$\mathsf{C}(x) = \begin{cases} \frac{x-1}{2} & 1 \leq x \leq 3 \\ \frac{5-x}{2} & 3 < x \leq 5 \end{cases} . \tag{7.7}$$

7.3.1.1. Interval arithmetic

How did we get the solution given in Eq. (7.7) for the fuzzy number $\mathsf{C}(x)$ that is the sum of the fuzzy numbers $\mathsf{A}(x)$ and $\mathsf{B}(x)$? It can be solved for directly by doing a bit of algebra, but an easier method is to use the interval interpretation of Kaufmann and Gupta (1985). Remember that a fuzzy set is completely characterized by its α–cuts. Since a fuzzy number is a convex fuzzy set defined on some subset of the real numbers, the alpha cuts are intervals,

$$\mathsf{A}^{\alpha} = [\underline{a}, \overline{a}] = \{x \mid \mathsf{A}(x) \geq \alpha\}$$

where

$$\underline{a} = \min\{x \mid \mathsf{A}(x) \geq \alpha\} \tag{7.8}$$

and

$$\overline{a} = \max\{x \mid \mathsf{A}(x) \geq \alpha\} \tag{7.9}$$

Remark 6. The notation $[\underline{a}, \overline{a}]$ comes from the branch of mathematics called interval analysis.

Therefore the set of pairs $\langle \alpha, \mathsf{A}^{\alpha} \rangle$, for every $\alpha \in [0,1]$, completely characterize a fuzzy number and consist of a level of presumption α and of an interval of confidence $\mathsf{A}^\alpha = [\underline{a}, \overline{a}]$.

Since A is a fuzzy number, at each presumption level α the interval of confidence is simply an interval on the real number line (if the domain is the integers or natural numbers instead of intervals one gets consecutive sequences but this makes little difference to the algebra involved.) In interval analysis the sum of two intervals $[\underline{a}, \overline{a}]$ and $[\underline{b}, \overline{b}]$ is simply another interval $[\underline{a}+\underline{b}, \overline{a}+\overline{b}]$,

$$[\underline{a}, \overline{a}] \oplus [\underline{b}, \overline{b}] = [\underline{a}+\underline{b}, \overline{a}+\overline{b}] \quad .$$

Similarly the product of two intervals is given by the formula

$$[\underline{a}, \overline{a}] \otimes [\underline{b}, \overline{b}] = [\underline{a} \cdot \underline{b}, \overline{a} \cdot \overline{b}]$$

the formula for the negative of an interval is

$$\ominus[\underline{a}, \overline{a}] = [-\overline{a}, -\underline{a}]$$

and the formula for the multiplicative inverse of an interval is

$$[a, \overline{a}]^{-1} = [\overline{a}^{-1}, \underline{a}^{-1}] \text{ if } \ \underline{a}, \ \overline{a} \geq 0.$$

This allows for a complete algebra of intervals as the subtraction of intervals is defined by the formula

$$[\underline{a}, \overline{a}] \ominus [\underline{b}, \overline{b}] = [\underline{a}, \overline{a}] + (\ominus[\underline{b}, \overline{b}])$$

and the quotient of two intervals is defined by

$$[\underline{a}, \overline{a}] \oslash [\underline{b}, \overline{b}] = [\underline{a}, \overline{a}] \otimes ([\underline{b}, \overline{b}]^{-1})$$

We can now show that if $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$ then $\mathsf{C}^\alpha = \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ for all α, where $\mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ is interval addition. Thus if we need to add two fuzzy numbers we can do it by adding the corresponding intervals of confidence (α-cuts) at each level of presumption (α-level).

Remark 7. All of these results are also basic results from interval analysis.

Theorem 9. *If* $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$ *then* $\mathsf{C}^\alpha = \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ *for all* $\alpha \in I$.

Proof. To prove $\mathsf{C}^\alpha = \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ we shall show that $\mathsf{C}^\alpha \subseteq \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ and $\mathsf{C}^\alpha \supseteq \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$.

To show that $\mathsf{C}^\alpha \subseteq \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ is true is completely straightforward. $\mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$, by the definition of interval addition, is the interval $[\underline{a}+\underline{b}, \overline{a}+\overline{b}]$, where the endpoints $\underline{a}$ and $\overline{a}$ are defined in Eqs. (7.8) and (7.9).

The values of and $\underline{b}$ and $\overline{b}$ are defined similarly. The interval $[\underline{a}+\underline{b}, \overline{a}+\overline{b}]$ contains an element z if and only if there is an x in $[\underline{a}, \overline{a}]$ and a y in $[\underline{b}, \overline{b}]$ such that $z = x + y$. But if that is true then C^α contains z since $\mathsf{A}(x) \geq \alpha$ and $\mathsf{B}(y) \geq \alpha$ which means that $\min[\mathsf{A}(x), \mathsf{B}(y)]$ is greater than or equal to α and hence z is an element of C^α . Therefore every element of $\mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$ is an element of C^α and $\mathsf{C}^\alpha \subseteq \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$.

Now suppose that $z \in \mathsf{C}^\alpha$. Then for some x and y such that $x + y = z$ the minimum of $\mathsf{A}(x)$ and $\mathsf{B}(y)$ was greater than or equal to α . But for the minimum of $\mathsf{A}(x)$ and $\mathsf{B}(y)$ to be greater than alpha both $\mathsf{A}(x)$ and $\mathsf{B}(y)$ must be greater than alpha. If $\mathsf{A}(x)$

is greater than α then x is an element of $\mathsf{A}^\alpha = [\underline{a}, \overline{a}]$ and, similarly, y is an element of $\mathsf{A}^\alpha = [\underline{b}, \overline{b}]$. Consequently $z = x + y$ must be an element of $\mathsf{A}^\alpha + \mathsf{B}^\alpha$ or $\mathsf{C}^\alpha \supseteq \mathsf{A}^\alpha \oplus \mathsf{B}^\alpha$. $\square$

A completely similar proof can show that if $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ then for every alpha, $\alpha \in [0, 1]$, that $\mathsf{C}^\alpha = \mathsf{A}^\alpha \otimes \mathsf{B}^\alpha$. Again we use $\otimes$ for fuzzy number multiplication to emphasize that it is not the product of the membership functions of A and B. The negative of a fuzzy number B, $\ominus$B, gives reverse negative intervals for each α. The reciprocal of a fuzzy number B, B^{-1}, gives reverse reciprocal intervals for each α.

Let us re-examine the problem introduced earlier in this chapter. We want to add the fuzzy numbers A and B with membership functions

$$\mathsf{A}(x) = \begin{cases} x & 0 \le x \le 1 \\ 2 - x & 1 < x \le 2 \end{cases}$$

and.

$$\mathsf{B}(x) = \begin{cases} x - 1 & 1 \le x \le 2 \\ 3 - x & 2 < x \le 3 \end{cases}$$

and produce a formula for their sum, product, difference and quotient.

To get a feel for the results and methods that follow let us first use some hard numbers. Let us fix alpha, $\alpha = 0.3$. Then the alpha–cuts of A and B are intervals, specifically $\mathsf{A}^{0.3} = [0.3, 1.7]$ and $B^{0.3} = [1.3, 2.7]$.

Addition When we add intervals we add the corresponding endpoints so that

$$\begin{aligned} \mathsf{A}^{0.3} \oplus \mathsf{B}^{0.3} &= [0.3, 1.7] \oplus [1.3, 2.7] \\ &= [1.6, 4.4] \end{aligned}$$

which must be the alpha-cut of $\mathsf{C} = \mathsf{A} \oplus \mathsf{A}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [1.6, 4.4]$.

Subtraction When we subtract intervals (with all endpoints positive numbers) we reverse the second interval and then subtract the corresponding endpoints so that

$$\begin{aligned} \mathsf{A}^{0.3} \ominus \mathsf{B}^{0.3} &= [0.3, 1.7] \ominus [1.3, 2.7] \\ &= [-2.4, 0.4] \end{aligned}$$

which must be the alpha-cut of $\mathsf{C} = \mathsf{A} \ominus \mathsf{B}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [-2.4, 0.4]$.

Multiplication When we multiply intervals we multiply the corresponding endpoints so that

$$\begin{aligned} \mathsf{A}^{0.3} \otimes \mathsf{B}^{0.3} &= [0.39, 1.7] \otimes [1.3, 2.7] \\ &= [0.09, 4.59] \end{aligned}$$

which must be the alpha-cut of $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [0.09, 4.59]$.

Division When we divide intervals (with all endpoints positive numbers) we reverse the second interval and then divide the corresponding endpoints so that

$$\begin{aligned} \mathsf{A} \oslash \mathsf{B}^{0.3} &= [0.3, 1.7] \oslash [1.3, 2.7] \\ &= [0.11, 1.31] \end{aligned}$$

to two decimal places, which must be the alpha-cut of $\mathsf{C} = \mathsf{A} \oslash \mathsf{B}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [0.11, 1.31]$.

Negation When we negate an intervals we reverse the endpoints and then negate them

$$\begin{aligned} \ominus \mathsf{B}^{0.3} &= \ominus [1.3, 2.7] \\ &= [-2.7, -1.3] \end{aligned}$$

which must be the alpha-cut of $\mathsf{C} = \mathsf{A} \ominus \mathsf{B}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [-2.7, -1.3]$.

Reciprocal When we negate an intervals we reverse the endpoints and then take their reciprocals

$$\begin{aligned} \left(\mathsf{B}^{0.3}\right)^{-1} &= [1.3, 2.7]^{-1} \\ &= [0.37, 0.77] \end{aligned}$$

to two decimal places, which must be the alpha-cut of $\mathsf{C} = \mathsf{B}^{-1}$ at $\alpha = 0.3$ or $\mathsf{C}^{0.3} = [0.37, 0.77]$.

Now a more general case with the fuzzy numbers A and B. First off, we can notice that A and B are triangular fuzzy numbers or TFs. This will keep the calculations simpler in the next set of examples. Now for each α the alpha-cut A^α is an interval, and because of the triangular nature of the membership function we can solve for the left-hand and right-hand limits of the alpha-cuts in terms of α. For an arbitrary α,$\mathsf{A}^\alpha = [\underline{a}, \overline{a}]$ we know that $\underline{a}$ and $\overline{a}$ depend on alpha, that is

$$\mathsf{A}^\alpha = [\underline{a}(\alpha), \overline{a}(\alpha)]$$

Also the left and right hand limits of the alpha-cut of B depend only on the value of α:

$$\mathsf{B}^\alpha = [\underline{b}(\alpha), \overline{b}(\alpha)]\,.$$

We can solve explicitly for these functions of alpha using Eqs. (7.1) and (7.2). For instance if we take the right hand side of the fuzzy triangular number B and set it equal to α we get

$$\alpha = 3 - x$$

and solving this for the x value produces $x = 3 - \alpha$ which will be the right hand endpoint of the alpha-cut of B. Similarly, setting α equal to the left hand side and solving produces $x = \alpha + 1$ which will be the left endpoint of the alpha-cut interval. Repeat this for the fuzzy number A and one gets the following formulas for the left and right endpoints of the intervals of the alpha-cuts of A and B:

$$\begin{aligned} \underline{a}(\alpha) &= \alpha, \\ \overline{a}(\alpha) &= 2 - \alpha. \\ \underline{b}(\alpha) &= \alpha + 1. \\ \overline{b}(\alpha) &= 3 - \alpha \quad . \end{aligned}$$

Another way of saying this is that for any $\alpha \in [0,1]$ the alpha-cuts are the intervals:

$$\mathsf{A}^\alpha = [\alpha, 2-\alpha] \text{ and}$$
$$\mathsf{B}^\alpha = [\alpha+1, 3-\alpha] \ .$$

An earlier theorem showed that if $\mathsf{C} = \mathsf{A} \oplus \mathsf{B}$ then the α-cuts of C were equal to the sum of the alpha-cuts of A and B. Therefore

$$\begin{aligned} \mathsf{C}^\alpha &= [\underline{c}(\alpha), \overline{c}(\alpha)] \\ &= [\underline{a}(\alpha), \overline{a}(\alpha)] \oplus [\underline{b}(\alpha), \overline{b}(\alpha)] \\ &= [\underline{a}(\alpha) + \underline{b}(\alpha), \overline{a}(\alpha) + \overline{b}(\alpha)] \end{aligned}$$

and if we replace $\underline{a}(\alpha)$ with α, $\overline{a}(\alpha)$ with $2-\alpha$, $\underline{b}(\alpha)$ with $\alpha+1$ and $\overline{b}(\alpha)$ with $3-\alpha$ this gives

$$\begin{aligned} \mathsf{C}^\alpha &= [\alpha, 2-\alpha] \oplus [\alpha+1, 3-\alpha] \\ &= [2\alpha+1, 5-2\alpha] \end{aligned}$$

We now know that the left end of the α-cut interval of C, in terms of α, is

$$\underline{c}(\alpha) = 2\alpha + 1 \tag{7.10}$$

and the right end of the α-cut interval of C, in terms of α, is

$$\overline{c}(\alpha) = 5 - 2\alpha. \tag{7.11}$$

Next, solve Eq. (7.10) for α in terms of $\underline{c}$ to give the left-hand function, ${}^LF$, for the fuzzy number C. Solving $\underline{c}(\alpha) = 2\alpha + 1$ for α gives

$$\alpha = \frac{\underline{c}(\alpha) - 1}{2} \ . \tag{7.12}$$

Then, solve Eq. (7.11) and for α in terms of $\bar{c}$ to get the right-hand function, ${}^RF$, for the $L-R$ fuzzy number C. Solving $\overline{c}(\alpha) = 5 - 2\alpha$ for α gives

$$\alpha = \frac{5 - \overline{c}(\alpha)}{2} \ . \tag{7.13}$$

It is important to understand here that α is essentially a membership grade and $\underline{c}(\alpha)$ is the leftmost x value at membership grade α and $\bar{c}$ is the rightmost x value at membership grade α. If this is understood properly, we realize that Eq. (7.12) is the equation of the left hand (ascending) side of the fuzzy number C and Eq. (7.13) is the equation of the right hand (descending) side of the fuzzy number C. Putting it all together, we get the following formula for the fuzzy number C:

$$\mathsf{C}(x) = \begin{cases} \frac{x-1}{2} & 1 \le x \le 3 \\ \frac{5-x}{2} & 3 < x \le 5 \end{cases}$$

The intervals of definition, for example $1 \le x \le 3$, can be determined by adding the corresponding intervals of definition of A, which is $0 \leq x \leq 1$, and B, which is $1 \leq x \leq 2$, or by determining where $\frac{x-1}{2}$ takes on values in the unit interval, $0 \leq \frac{x-1}{2} < 1$. Both methods give the same result. This number is pictured in Fig. (7.11).As a second example let us find the formulas for the alpha-cuts of the fuzzy number C that is the difference of A and B. With $\mathsf{C} = \mathsf{A} \ominus \mathsf{B}$

$$\begin{aligned} \mathsf{C}^{\alpha} &= [\underline{c}(\alpha), \overline{c}(\alpha)] \\ &= [\underline{a}(\alpha), \overline{a}(\alpha)] \ominus [\underline{b}(\alpha), \overline{b}(\alpha)] \\ &= [\underline{a}(\alpha) - \overline{b}(\alpha), \overline{a}(\alpha) - \underline{b}(\alpha)] \\ &= [\alpha, 2-\alpha] \ominus [\alpha+1, 3-\alpha] \\ &= [\alpha - (3-\alpha), (2-\alpha) - (\alpha+1)] \\ &= [2\alpha - 3, 1 - 2\alpha] \end{aligned}$$

We now know that for subtraction

$$\underline{c}(\alpha) = 2\alpha - 3 \tag{7.14}$$

and

$$\overline{c}(\alpha) = 1 - 2\alpha. \tag{7.15}$$

Solving these equations for $\underline{c}$

$$\alpha = \frac{\underline{c}(\alpha) + 3}{2} \ . \tag{7.16}$$

and $\overline{c}$

$$\alpha = \frac{1 - \overline{c}(\alpha)}{2} \ . \tag{7.17}$$

respectively, gives us the shape of the fuzzy number $\mathsf{C} = \mathsf{A} \ominus \mathsf{B}$:

$$\mathsf{C}(x) = \begin{cases} \frac{x+3}{2} & -3 \leq x \leq -1 \\ \frac{1-x}{2} & -1 < x \leq 1 \end{cases}$$

The fuzzy number C is a triangular fuzzy number, $\mathsf{C} = \mathsf{Tr}[-3, -1, 1]$.

The following examples show that the product and quotient of triangular fuzzy numbers are not triangular fuzzy numbers.For the next example let us find the formulas for the alpha-cuts of the fuzzy number C that is the product of A and B. With $\mathsf{C} = \mathsf{A} \otimes \mathsf{B}$ we deduce

$$\begin{aligned} \mathsf{C}^{\alpha} &= [\underline{c}(\alpha), \overline{c}(\alpha)] \\ &= \mathsf{A}^{\alpha} \otimes \mathsf{B}^{\alpha} \\ &= [\underline{a}(\alpha), \overline{a}(\alpha)] \otimes [\underline{b}(\alpha), \overline{b}(\alpha)] \\ &= [\underline{a}(\alpha) \cdot \underline{b}(\alpha), \overline{a}(\alpha) \cdot \overline{b}(\alpha)] \\ &= [\alpha \cdot (\alpha+1), (3-\alpha) \cdot (2-\alpha)] \\ &= [\alpha^2 + \alpha, 6 - 5\alpha + \alpha^2] \end{aligned}$$

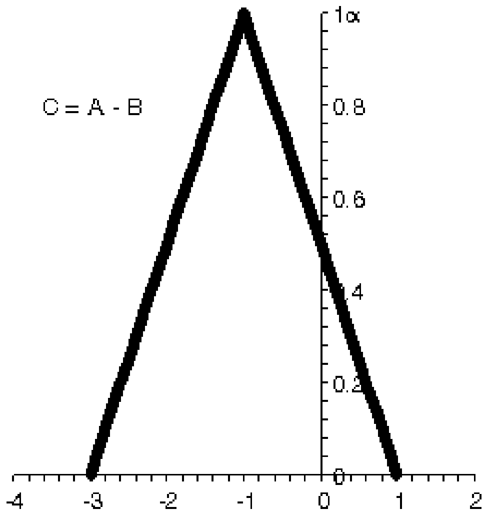


Figure 7.12.: The fuzzy number $C = A - B$.

So the left end of the α-cut interval, in terms of α, is

$$\underline{c}(\alpha) = \alpha^2 + \alpha$$

and the right end of the α-cut interval, in terms of α, is

$$\overline{c}(\alpha) = 6 - 5\alpha + \alpha^2.$$

We can now solve these two equations for α in terms of $\underline{c}$ to give the left-hand function, ${}^{L}F$, for the fuzzy number C and for α in terms of $\overline{c}$ to get the right-hand function, ${}^{R}F$, for the $L - R$ fuzzy number C using the quadratic equation

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and keeping only the positive roots for the appropriate intervals. The final result is

$$\mathsf{C}(x) = \begin{cases} \frac{\sqrt{4x+1}-1}{2} & 0 \leq x \leq 2 \\ \frac{5-\sqrt{4x+1}}{2} & 2 < x \leq 6 \end{cases}.$$

The last example finds the formulas for the alpha-cuts of the fuzzy number C that is

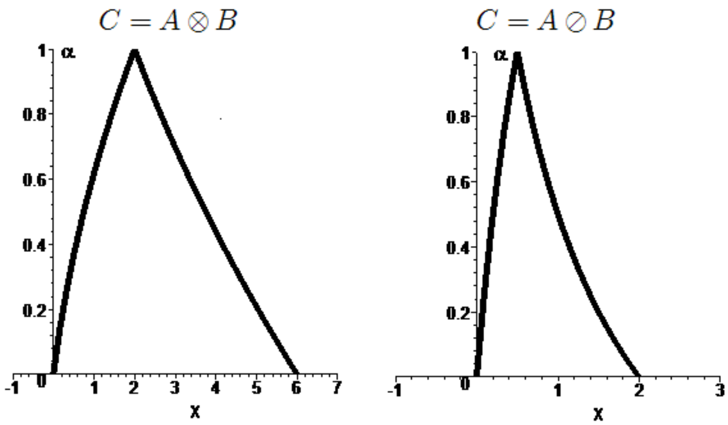


Figure 7.13.: Fuzzy arithmetic: the product and quotient of A and B, $A \otimes B$ and $A \oslash B$.

the quotient of A and B if B has positive values for all its alpha–cuts. With $\mathsf{C} = \mathsf{A} \oslash \mathsf{B}$

$$\begin{aligned}
\mathsf{C}^\alpha &= [\underline{c}(\alpha), \overline{c}(\alpha)] \\
&= \mathsf{A}^\alpha \oslash \mathsf{A}^\alpha \\
&= [\underline{a}(\alpha), \overline{a}(\alpha)] \oslash [\underline{b}(\alpha), \overline{b}(\alpha)] \\
&= \left[\frac{\underline{a}(\alpha)}{\overline{b}(\alpha)}, \frac{\overline{a}(\alpha)}{\underline{b}(\alpha)}\right] \\
&= [\alpha, 2-\alpha] \oslash [\alpha+1, 3-\alpha] \\
&= \left[\frac{\alpha}{3-\alpha}, \frac{2-\alpha}{\alpha+1}\right]
\end{aligned}$$

So

$$\underline{c}(\alpha) = \frac{\alpha}{3-\alpha}$$

and

$$\overline{c}(\alpha) = \frac{2-\alpha}{\alpha+1}.$$

We can now solve these for α in terms of $\underline{c}$ to give the left-hand function, ${}^LF$, for the fuzzy number C and for α in terms of $\overline{c}$ to get the right-hand function, ${}^RF$, for the $L-R$ fuzzy number C. A little algebra gives

$$\mathsf{C}(x) = \begin{cases} \frac{3x}{1+x} & 0 \le x \le \frac{1}{2} \\ \frac{2-x}{x+1} & \frac{1}{2} < x \le 2 \end{cases} .$$

There is one problem that should be mentioned in connection with reliance on interval based methods in computing functions of fuzzy numbers, this is the cancellation problem. Consider the functions $f(x) = x - x$ and $g(x) = x \div x$. If we use $\mathsf{A}(x)$, almost one, and blindly do interval arithmetic we get that $\mathsf{C} = f(\mathsf{A}) = Tr\,\langle 0, 2, 4\rangle$ however if we use the extension principle we get an impulse number $\mathsf{C}' = \mathrm{Tr}\,\langle 0, 0, 0\rangle$ since $x - x$ is always zero! See Klir (1997) for a complete analysis of the difference between interval and extension methods.

7.4. Fuzzy functions

A normal function, such as $f(x) = x^2$ maps numbers to numbers. The function $f(x) = x^2$ would map the number $x = 3$ to $f(3) = 9$. A fuzzy function should map a fuzzy number to a fuzzy number.

Fuzzy numbers Let $X \neq \emptyset$ and $Y \neq \emptyset$ be universal crisp sets and let f be a function from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$. Then f is called a fuzzy function (or mapping) and we use the notation $f : \mathcal{F}(X) \to \mathcal{F}(Y)$.

It should be noted, however, that a fuzzy function is not necessarily defined by Zadeh's extension principle. It can be any function which maps a fuzzy set $\mathsf{A} \in \mathcal{F}(X)$ into a fuzzy set $\mathsf{B} \equiv f(\mathsf{A}) \in \mathcal{F}(Y)$.

Definition 21. Let $X \neq \emptyset$ and $Y \neq \emptyset$ be crisp sets. A fuzzy mapping $f : \mathcal{F}(X) \to \mathcal{F}(Y)$ is said to be monotone increasing if for every $\mathsf{A}, \mathsf{A}' \in \mathcal{F}(X)$ and $\mathsf{A} \subset \mathsf{A}'$ it follows that $f(\mathsf{A}) \subset f(\mathsf{A}')$.

Theorem 10. *Let $X \neq \emptyset$ and $Y \neq \emptyset$ be crisp sets. Then every fuzzy mapping $f : \mathcal{F}(X) \to \mathcal{F}(Y)$ defined by the extension principle is monotone increasing.*

Proof. Let $\mathsf{A}, \mathsf{A}' \in \mathcal{F}(X)$ be such that $\mathsf{A} \subset \mathsf{A}'$. Then using the definition of extension principle we get

$$f(\mathsf{A})(y) = \sup_{x \in f^{-1}(y)} \mathsf{A}(x) \leq \sup_{x \in f^{-1}(y)} \mathsf{A}'(x) = f(\mathsf{A}')(y)$$

for all $y \in Y$. □

Lemma 1. *Let $\mathsf{A}, \mathsf{B} \in \mathcal{F}$ be fuzzy numbers and let $f(\mathsf{A}, \mathsf{B}) = \mathsf{A} + \mathsf{B}$ be defined by the extension principle. Then f is monotone increasing.*

Proof. Let $\mathsf{A}, \mathsf{A}', \mathsf{B}, \mathsf{B}' \in \mathcal{F}$ such that $\mathsf{A} \subset \mathsf{A}'$ and $\mathsf{B} \subset \mathsf{B}'$. Then using the definition of extension principle we get

$$(\mathsf{A} + \mathsf{B})(z) = \sup_{x+y=z} \min [\mathsf{A}(x), \mathsf{B}(y)] \leq \sup\ _{x+y=z} \min [\mathsf{A}'(x), \mathsf{B}'(y)]\} = (\mathsf{A}' + \mathsf{B}')(z)$$

□

Lemma 2. *Let $\mathsf{A}, \mathsf{B} \in \mathcal{F}$ be fuzzy numbers, let λ_1, λ_2 be real numbers and let $f(\mathsf{A}, \mathsf{B}) = \lambda_1 \mathsf{A} + \lambda_2 \mathsf{B}$ be defined by the extension principle. Then f is a monotone increasing fuzzy function.*

The above results can be generalized to linear combinations of fuzzy numbers.

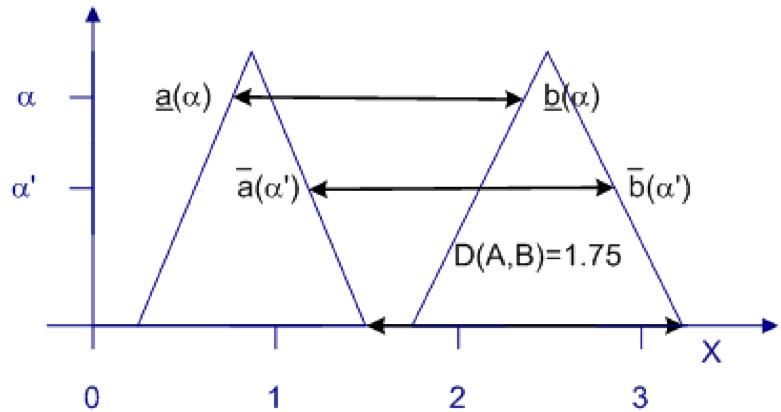


Figure 7.14.: Distance between fuzzy numbers A and B using the distance D.

7.5. Metrics for fuzzy numbers

When we are given two numbers, like $a = 12$ and $b = 7$, it is quite easy to say how far apart they are, or, technically, the *distance* between a and b. They are 5 units apart and the formula for the distance is the absolute value of the difference:

$$dist(a, b) = |a - b| \ .$$

For fuzzy numbers a more complex definition is needed, in fact more than one type of distance has been proposed. The ones that are applied the most often are the Hausdorf, C_∞, Hamming, and Discrete Hamming distances. See 2.6 on page 18.

In all of the following definitions A and B are fuzzy numbers, the α–cuts of A and B are $\mathsf{A}^\alpha = [\underline{a}(\alpha), \overline{a}(\alpha)]$ and $\mathsf{B}^\alpha = \left[\underline{b}(\alpha), \overline{b}(\alpha)\right]$ and the membership functions are $\mathsf{A}(x)$ and $\mathsf{B}(x)$.

Hausdorf distance The Hausdorf distance $Dist_D(\mathsf{A}, \mathsf{B})$ is the maximal distance between α level sets of A and B

C_∞ distance The C_∞ distance $C_\infty(\mathsf{A}, \mathsf{B})$ is the maximal distance between the membership grades of A and B

$$Dist_{C_\infty}(\mathsf{A}, \mathsf{B}) = \|\mathsf{A} - \mathsf{B}\|_\infty = \sup_{x \in X} |\mathsf{A}(x) - \mathsf{B}(x)| \, .$$

Hamming distance The Hamming distance $H(\mathsf{A}, \mathsf{B})$ is the integral of the distance between the membership grades of A and B

$$Dist_H(\mathsf{A}, \mathsf{B}) = \int_X |\mathsf{A}(x) - \mathsf{B}(x)| \ dx.$$

Discrete Hamming distance The Discrete Hamming distance $H(\mathsf{A}, \mathsf{B})$ is the sum of the distance between the membership grades of A and B for each of the n elements

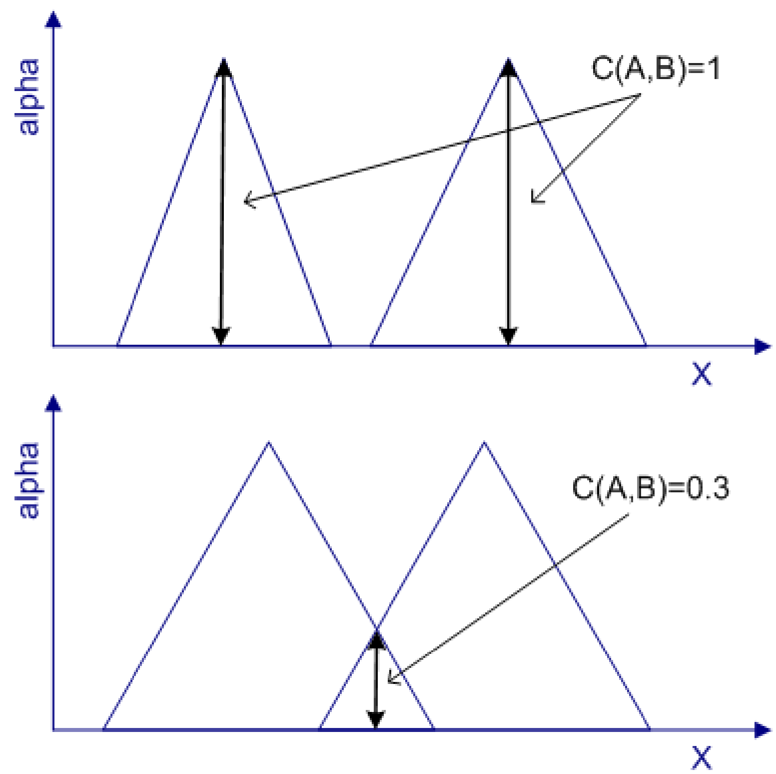


Figure 7.15.: Distance between fuzzy numbers A and B using distance C.

of the discrete universe $X = \{x_1, x_2, ..., x_n\}$

$$Dist_H(\mathsf{A}, \mathsf{B}) = \sum_{i=1}^{n} |\mathsf{A}(x_i) - \mathsf{B}(x_i)| \ .$$

7.6. Fuzzy Algebra

Now this book focuses on application and not theory, and the following short section can be omitted without consequence. In fact, if the reader does not have a background that includes limits and convexity the following proof would be difficult to follow. It is included to show the necessity of the rather complex definition of *fuzzy number,* specifically the convexity and continuity conditions.

Theorem 11. *The sum of two fuzzy numbers is a fuzzy number.*

Proof. Let m and n be arbitrary fuzzy numbers. Thus they are both normal, convex and upper semi-continuous functions with domain the real numbers. Equation (7.6) involves only the min and sup function. The min of two values in the unit interval is also in the unit interval. The sup of a sequence of values in the unit interval is in the unit interval. Thus $m \oplus n$ is a fuzzy set defined upon the real numbers. Since there exists $x, y \in \mathbb{R}$ such that $m(x) = n(y) = 1$ (they are normal) then we can conclude that $m \oplus n\,(x + y) = 1$ since the min of $m(x)$ and $n(y)$ is one and this value must be the supremum (there can be no larger membership values).

Let the α–cuts of m and n be $[\underline{m}, \overline{m}]$ and $[\underline{n}, \overline{n}]$. The α–cut of $m + n$ is $[\underline{m} + \underline{n}, \overline{m} + \overline{n}]$ and since $\underline{m} \leq \overline{m}$ and $\underline{n} \leq \overline{n}$ we have that $\underline{m} + \underline{n} \leq \overline{m} + \overline{n}$ and thus the α–cut of $m + n$ is a closed interval for arbitrary α. Thus $m + n$ is convex since all of its α–cuts are convex.

A function is upper semi-continuos if $\lim_{x \to a^+} f(x) = f(a)$. Since the sum of limits is the limit of the sum $\lim_{x \to a^+} [m + n]\,(x) = \lim_{x \to a^+} m(x) + \lim_{x \to a^+} n(x) = m(a) + n(a)$. □

The addition and multiplication of fuzzy numbers is commutative and associative:

$$a \oplus b = b \oplus a \tag{7.18}$$

$$(a \oplus b) \oplus c = a \oplus (b \oplus c) \tag{7.19}$$

and

$$a \otimes b = b \otimes a \tag{7.20}$$

$$(a \otimes b) \otimes c = a \otimes (b \otimes c) \tag{7.21}$$

7.7. Notes

Interval arithmetic is thoroughly covered in Moore (1963), Moore (1979), and Moore (1988). Fuzzy numbers and fuzzy arithmetic are best covered in Kaufmann and Gupta (1985) and Dubois and Prade (1987).

7.8. Homework

Let

$$\mathsf{A}(x) = \begin{cases} x & x \in [0,1] \\ 2 - x & x \in (1,2] \end{cases}$$

$$\mathsf{B}(x) = \begin{cases} x - 1 & x \in [1,2] \\ 3 - x & x \in (2,3] \end{cases}$$

$$\mathsf{C}(x) = \begin{cases} \frac{x-2}{2} & x \in [2,4] \\ \frac{6-x}{2} & x \in (4,6] \end{cases}$$

and

$$\mathsf{D}(x) = \begin{cases} \frac{x}{3} & x \in [0,3] \\ 1 & x \in (3,4] \\ 5 - x & x \in (4,5] \end{cases} \tag{7.22}$$

1. Graph A, B, C, and D.
2. Express A—D as fuzzy triangular or trapezoidal numbers, as appropriate.
3. What is $A \oplus B$.

4. What is $B \oplus A$.
5. What is $A \oplus C$.
6. What is $B \oplus C$.
7. What is $B \ominus A$.
8. What is $A \ominus B$.
9. What is $A \otimes B$.
10. What is $B \oslash A$.
11. What is $A \oslash B$.
12. What is C^{-1}.
13. What is $A \otimes (B \oplus C)$.
14. What is $(A \otimes B) \oplus (A \otimes C)$.
15. Given the above two results, do you think that multiplication distributes over addition?
16. Is $A \cap B$ a fuzzy number? Why?
17. Is $A \cup B$ a fuzzy number? Why?
18. Is $(A \oplus B) \cup D$ a fuzzy number? Why?
19. Express $A \oplus B$, $B \ominus A$, $A \otimes B$, and $B \oslash A$ as fuzzy triangular or trapezoidal numbers, if this is possible.
20. Show that $Tr\langle a, b, c\rangle \oplus Tr\langle d, e, f\rangle$ is $Tr\langle a+d, b+e, c+f\rangle$.
21. Show that $Tr\langle a, b, c\rangle \ominus Tr\langle d, e, f\rangle$ is $Tr\langle a-f, b-e, c-d\rangle$ if a—f are all nonnegative.
22. Derive a formula for $Tr\langle a, b, c\rangle \otimes Tr\langle d, e, f\rangle$ if a—f are all nonnegative.
23. Derive a formula for $Tr\langle a, b, c\rangle \oslash Tr\langle d, e, f\rangle$ if a—f are all nonnegative.
24. How would one represent a fuzzy number in C or C++ code.

8. Relations

8.1. Introduction

Relations are an important concept in mathematics and in the real world. A family is a group of people who have some relation with each other. Consider the set of people $\{\mathrm{Eve}, \mathrm{Fred}, \mathrm{Mary}, \mathrm{Sally},$ Tom,Bob$\}$ and the relation *father of*. Bob is the *father of* Sally. Tom is the *father of* Mary. Next consider the relation *parent of*. Bob is the *parent of* Sally. Sally is the *parent of* Mary. From all this information we might be able to conclude that Bob is the *grandfather of* Mary and that Tom is the *husband of* Sally.

There are many other types of relations in this world, relations between employers and employees, and relations between friends.

Friendship is certainly a very different kind of relationship than parenthood. It is basically true that adult P is the parent of child C or they are not. Friendship, on the other hand is not so cut and dried. There are friends who you can count on and there are friends you just met. There are friends on the way in and friends who are definitely on the way out.

While crisp relationships are perfect for encompassing ideas like *parent of,* fuzzy sets are much better at capturing relationships of degree, like friendship.

In crisp mathematics, equality is an important relationship and things are either equal or they are not. In fuzzy set theory, similarity is an important relationship, and things are similar to a degree. Similarity is a very context dependant concept, a house-cat and a tiger are *similar in shape* but not in size. A house-cat and toaster are *similar in size* but not in shape. Similarity is a very fuzzy concept.

8.2. Classical relations

A classical binary relation can be considered as a set of ordered pairs derved from a product space $X \times Y$. For example, if $X = \{a, b, c\}$ and $Y = \{1, 2, 3\}$ then a relation between X and Y is made explicit by indicating what elements of X have a specific relation, say R, with elements of Y. As noted in Sec. (2.4) there are many ways to present a relation. We can list the ordered pairs that are in the relation using set notation;

$$R = \{\langle a, 1\rangle, \langle a, 3\rangle, \langle b, 2\rangle, \langle c, 2\rangle, \langle c, 3\rangle\} \tag{8.1}$$

or use an in-line notation;

$$a\,R\,1, \quad a\,R\,3, \quad b\,R\,2, \quad c\,R\,2, \quad c\,R\,3 \tag{8.2}$$

R	1	2	3
a	1	0	1
b	0	1	0
c	0	1	1

Table 8.1.: A binary relation R presented as a table.

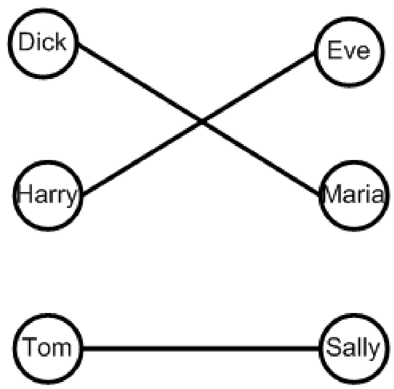


Figure 8.1.: Married.

We can also use a characteristic function to denote a relation, since a relation is just a set (of ordered pairs);

$$\chi_R(x,y) = \begin{cases} 1 & \langle x,y\rangle \in R \\ 0 & \text{otherwise} \end{cases} \tag{8.3}$$

8.2.1. Types of classical relations

Definition 22 (Binary Relation)**.** A binary relation between the sets X and Y is a subset of $X \times Y$. If R is a subset of $X \times Y$ and if $\langle x,y\rangle \in R$ then we say that x is related to y. If the sets X and Y are the same set, $X = Y$, then a subset R of $X \times X$ is called a relation on X.

By far the simplest way to present a binary relation R is in a table. In Table (8.1) a 1 in row a column 3 indicates that a is related to 3, or $a\,R\,3$, and the 0 in row b column 1 indicates that b is not related to 1, or $b \not R 1$.

Example 46. Let X be the domain of men $\{Tom,\ Dick,\ Harry\}$ and Y the domain of women $\{Eve, Maria, Sally\}$ then the relation "married to" on $X \times Y$ is, for example $\langle Tom, Sally\rangle$, $\langle Dick, Maria\rangle$, $\langle Harry, Eve\rangle$. See Fig. (8.1).

Binary relations are by far the most common relations one is likely to encounter. However there are some places, such as the creation of relational databases, that one encounters n-ary relations. In this case the relation is a subset of the product of n non-empty domain sets, $R \in X_1 \times X_2 \times \cdots \times X_n$. In this case the above notation (except for the in-line notation) can be generalized to higher dimensions.

Definition 23 (classical n-ary relation)**.** Let $X_1, \ldots, X_n$ be crisp sets. The subsets of the Cartesian product $\times_{i=1}^{n} X_i = X_1 \times \cdots \times X_n$ are called $n-$ary relations. If $X_1 = \cdots = X_n$ and $R \subseteq X^n$ then R is called an $n-$ary relation on X .

Example 47. Let S be the set of social security numbers, F be the set of first names, L be the set of last names, and G be the set of real numers between zero and five. A registrar's database would contain elements of $S \times F \times L \times G$ which allows the lookup of a students grade point average, or GPA. A typical element might be

$$\langle 123 - 45 - 6789, \text{Adam}, \text{Ant}, 2.3 \rangle \tag{8.4}$$

There are many different types of relations in mathematics. Three of the most important are identity, equivalence and order. The familiar equal sign, $=$, of mathematics is the most common identity relation. Identity is a special case of equivalence. An example of an equivalence relation on the natural numbers is; a is equivalent to b, written $a \approx b$, if and only if both a and b are even or both a and b are odd. Thus $2 \approx 4$ and $1 \approx 3$ but $2 \not\approx 3$. The relation less than, written $<$, is an order relation, as is the subset relation, $\subset$.

Every function $f : X \to Y$ can be considered as a relation. We can view f as the set of ordered pairs $f = \langle x, f(x) \rangle$. But a set of ordered pairs is precisely the technical definition of a relation.

Let R be a binary relation on a classical set X . For all of the following definitions, we assume that $x, y, z \in X$.

Definition (reflexive)**.** R is reflexive if $R(x, x)$, $\forall x \in X$.

Definition 24 (antireflexive)**.** R is antireflexive if not $R(x, x)$, $\forall x \in X$.

Definition 25 (symmetric)**.** R is symmetric when $R(x, y)$ if and only if $R(y, x)$.

Definition 26 (antisymmetric)**.** R is antisymmetric if $R(x, y)$ and $R(y, x)$ imply that $x = y$.

Definition 27 (transitive)**.** R is transitive if $R(x, y)$ and $R(y, z)$ imply that $R(x, z)$.

Example. Consider the classical identity, inequality and order relations on the real line. It is clear that equality, $=$, is reflexive, symmetric and transitive. Inequality, $\neq$, is antireflexive and symmetric. Less than, $<$, an order relation, is anti-reflexive, anti-symmetric and transitive.

Some of the more important classes of binary relations are:

Definition 28 (equivalence)**.** R is an equivalence relation if, R is reflexive, symmetric and transitive.

Often an equivalence relation is denoted by the symbol $\equiv$ which is used in-line, $x \equiv y$.

Definition 29 (quasi-equivalence)**.** R is a quasi-equivalence relation if, R is symmetric and transitive.

Definition 30 (compatibility)**.** R is a compatibility relation if, R is reflexive and symmetric. This type of relation is also called a *tolerance* relation.

	Reflexive	Antireflexive	Symmetric	Antisymmetric	Transitive
Equivalence	✓		✓		✓
Quasi-equivalence			✓		✓
Compatibility	✓		✓		
Partial order	✓			✓	✓
Preordering	✓				✓
Strict order		✓		✓	✓

Table 8.2.: Types of relations.

Definition 31 (partial order)**.** R is a partial order relation if it is reflexive, antisymmetric, and transitive.

Often a partial order is denoted by the symbol $\prec$ which is used in-line, $x \prec y$.

Definition 32 (preordering)**.** R is a preorder relation if it is reflexive and transitive.

Definition 33 (total order)**.** R is a total order relation if it is partial order and $\forall x, y \in X$, either $R(x, y)$ or $R(y, x)$ holds.

The following table summarizes the above definitions.

Example 48. Consider the relation a "the absolute difference of two numbers is divisible by three" or *mod 3* as it is termed in mathematics and computer science. This is a relation on the natural numbers $\mathbb{N}$. Thus $n \equiv m \mod 3$ if the difference, $m - n$, is exactly multiple of 3. Equivalently $n \equiv m \mod 3$ if the remainder upon division by three of m is the same as the remainder upon division by three of n. The $\operatorname{mod} 3$ relation is an equivalence relation.

Example 49. For any positive integer m the relation "the difference of a and b is divisible by m" is an equivalence relation upon any subset of the integers.

Example 50. *Subset of* is a partial order. It is not a total order because if $A = \{1, 2\}$ and $B = \{2, 3\}$ then neither A nor B is a subset of the other.

Example 51. Two elements of $\mathbb{N}_5$ are compatible, represented with $\sim$, if they are within a unit of each other. Thus $2 \approx 3$, and $3 \approx 4$ since $|2 - 3| \leq 1$ and $|3 - 4| \leq 1$. But $2 \not\approx 4$ so the comapatible relation is not transitive.

Example 52. Let $X = \mathbb{N}_5 \cup \{a\}$ and let the relation $\sim$ be numerical equality. Then $2 \sim 2$, $3 \sim 3$, etc., but $\sim$ is only a quasi-order as nothing in X is related to a.

Example 53. Let X be the set of all people and let the relation $\sim$ be identity or is descended from. This relation is reflexive since identity is reflexive and I am I. The relation is transitive since my descendant's descendant is also my descendant. But it is not symmetric as I am not descended from my children.

An important mathematical point is that every equivalence relation upon a set X partitions that set into pieces called equivalence classes. The mod 3 equivalence relation divides the natural number into three pieces $[1] = \{1, 4, 7, 10, \ldots\}$, $[2] = \{2, 5, 8, 11, \ldots\}$, $[3] = \{3, 6, 9, 12, \ldots\}$. This is a partition since none of the three pieces; $[1]$, $[2]$, $[3]$ contain any elements in common and their union is the domain set $\mathbb{N}$.

8.3. Fuzzy relations

If we examine the characteristic function view of a relation, Eq. (8.3) or for that matter the table version of Table (8.1) we see a prime example of a concept (the relation) that in traditional mathematics maps to a range set $\{0, 1\}$. We can easily fuzzify the concept of a relation by substituting the unit interval $[0, 1]$ as the range set of the membership function (which used to be called a characteristic function) of a relation R.

Definition 34 (fuzzy relation)**.** Let X and Y be nonempty sets. A fuzzy relation R is a fuzzy subset of $X \times Y$. In other words, $R \in \mathcal{F}(X \times Y)$. If $X = Y$ then we say that R is a binary fuzzy relation on X .

Let R be a binary fuzzy relation on X . Then $R(x, y)$ is interpreted as the degree of membership of the ordered pair (x, y) in R .

Example 54. A simple example of a binary fuzzy relation on $X = \{1, 2, 3\}$, called “approximately equal” can be defined as

$$R(1,1) = R(2,2) = R(3,3) = 1, \tag{8.5}$$

$$R(1,2) = R(2,1) = R(2,3) = R(3,2) = 0.7, \tag{8.6}$$

$$R(1,3) = R(3,1) = 0.4\,. \tag{8.7}$$

In other words, $R\,(x, y) = 1$ if $x = y$, 0.7 if $|x - y| = 1$, 0.4 if $|x - y| = 2$. In matrix notation the relation R can be represented as

R	1	2	3
1	1.0	0.7	0.4
2	0.7	1.0	0.7
3	0.4	0.7	1.0

or, even more abstractly, as

$$\begin{bmatrix} 1 & 0.7 & 0.4 \\ 0.7 & 1 & 0.7 \\ 0.4 & 0.7 & 1 \end{bmatrix}$$

where in the second matrix the X corresponds to rows and Y corresponds to columns.

8.3.1. Composition

If x is somewhat related to y and y also has some degree of relation to z then it makes sense that there is implicitly some degree of relationship between x and z.

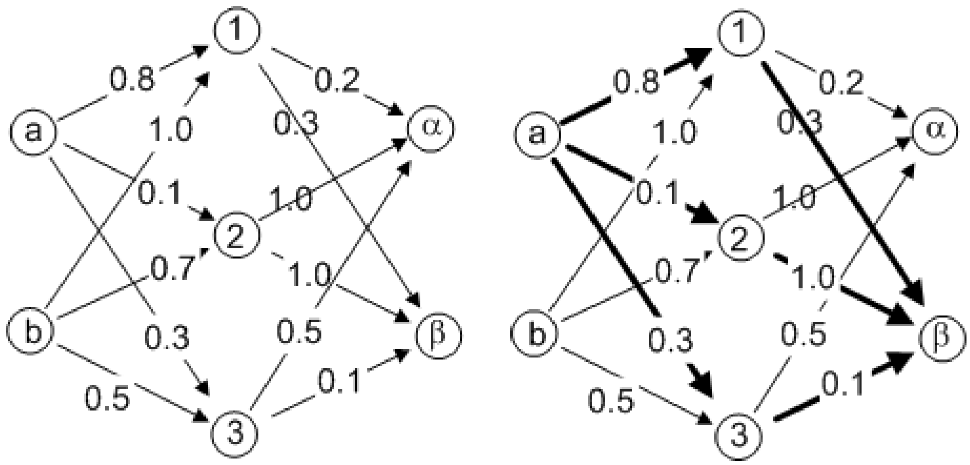


Figure 8.2.: A graph of the relations R and S.

Let R be a fuzzy relationship between X and Y and let S be a fuzzy relationship between Y and Z. Fuzzy set theory uses the principle that a chain is as strong as its weakest individual link and that a rope of chains is as strong as the strongest individual chain. This principle says that the degree of relationship between x and z through a single intermediate value y is the $\min$ of the membership grades of $\langle x, y\rangle \in R$ and $\langle y, z\rangle \in S$ (this is the weakest link). But since there may be many elements of Y through which x and z could be connected by chains the strength of the resultant link is the $\max$ of all the chains.

$$R \circ S\,(x, z) = \max_{y \in Y} \min[R(x, y), S(y, z)].$$

Let $X = \{a, b\}$, $Y = \{1, 2, 3\}$, $Z = \{\alpha, \beta\}$, and let $R : X \times Y \to [0, 1]$ and $S : Y \times Z \to [0, 1]$ be given by the following tables

R	1	2	3
a	0.8	0.1	0.3
b	1.0	0.7	0.5

S	α	β
1	0.2	0.3
2	1.0	1.0
3	0.5	0.1

(8.8)

This situation is illustrated in Fig. (8.2).

There is a technical detail that must be mentioned here about the real numbers. Some groups of real numbers do not contain a maximum. The set of negative real numbers does no contain a maximum value. Instead the negative numbers have a supremum or $\sup$ of zero, but zero is not a negative number, so zero is not in the set of negative numbers. Thus the set of negative numbers does not contain its supremeum. This is one of those properties of the real numbers that make calculus difficult and leads to limits and other troublesome concepts. Because of this, many books use max–min instead of sup–min since the term max–min is easier to understand. Technically, if we want to be mathematically correct we should have the following definition.

Definition 35 (sup–min composition)**.** Let $R \in \mathcal{F}\,(X \times Y)$ and $S \in \mathcal{F}(Y \times Z)$. The $\sup$–min composition of R and S, denoted by $R \circ S$ is defined as

$$(R \circ S)(x, z) = \sup_{y \in Y} \min\,[R(x, y), S(y, z)]$$

Example 55. Let $X = \{a, b\}$, $Y = \{1, 2, 3\}$, $Z = \{\alpha, \beta\}$, and let $R : X \times Y \to [0, 1]$ and $S : Y \times Z \to [0, 1]$ are given in Eq. (8.8) then the sup–min composition of R and S, $R \circ S$, is given by the table

$R \circ S$	α	β
a	0.3	0.3
b	0.7	0.7

In the second graph in (8.2) the three paths from a to β have been emphasized. Each path has two pieces, for example the edge $\alpha \to 1$ and the edge $1 \to \beta$ form one path from a to β. The edge weights are 0.8 and 0.3 respectively. If this path is stressed then it will overload when the stress exceeds 0.3 since then the edge $1 \to \beta$ will break. Thus each path is only as strong as its weakest link. There are three paths from a to β, one each going through nodes 1, 2, and 3. We have already seen that the path through node 1 breaks at stress level 0.3. The path through node 2 ($\alpha \to 2$, strength 0.1, and $2 \to \beta$, strength 0.3) breaks at stress level 0.1. The path through node 3 ($\alpha \to 3$, strength 0.3, and $3 \to \beta$, strength 0.1) also breaks at stress level 0.1. So when the stress hits 0.3 all three paths from a toβ are broken and that must be the strength of the relation between a and β in $R \circ S$.

Example 56. Consider two fuzzy relations

$R =$ “x is considerable larger than y$''$ =

R	y_1	y_2	y_3
x_1	0.8	0.1	0.1
x_2	0	0.8	0
x_3	0.9	1	0.7
x_4	0.8	0	0.7

and

$S =$ “y is very close to z$''$ =

S	z_1	z_2	z_3	z_4
y_1	0.4	0	0.9	0.6
y_2	0.9	0.4	0.5	0.7
y_3	0.3	0	0.8	0.5

.

Then their composition is

$R \circ S =$

$R \circ S$	z_1	z_2	z_3	z_4
y_1	0.4	0	0.9	0.6
y_2	0.9	0.4	0.5	0.7
y_3	0.3	0	0.8	0.5

Remark 8. The composition of R and S is performed very similarly to the classical product of the matrices R and S. The difference is that instead of multiplication we use the minimum operator and instead of addition we use the maximum operator.

We can also compose a fuzzy set with a relation, to produce a to produce a new fuzzy set.

Definition 36. If C is a fuzzy set on X and R is a fuzzy relation on $X \times Y$ then their composition is

$$C \circ R\,(y) = \sup_{x \in X} \min[C(x), R(x,y)]. \tag{8.9}$$

Definition 37. If R is a fuzzy relation on $X \times Y$ and D a fuzzy set on Y then their composition is

$$R \circ D\,(y) = \sup_{y \in Y} \min[R(x,y), D(y)]. \tag{8.10}$$

Example 57. Let C be a fuzzy set in the universe of discourse $\{1,2,3\}$ and let R be a binary fuzzy relation on $\{1,2,3\}$. Assume that $C = \{\langle 1, 0.2\rangle, \langle 2, 1.0\rangle, \langle 3, 0.2\rangle\}$
and that R is giving in the following table

R	1	2	3
1	1	0.8	0.3
2	0.8	1	0.8
3	0.3	0.8	1

. (8.11)

Using the definition of $\sup - \min$ composition we get

$$C \circ R = \begin{bmatrix} 0.2 & 1 & 0.2 \end{bmatrix} \circ \begin{bmatrix} 1 & 0.8 & 0.3 \\ 0.8 & 1 & 0.8 \\ 0.3 & 0.8 & 1 \end{bmatrix} = \begin{bmatrix} 0.8 & 1 & 0.8 \end{bmatrix}$$

so

$$C \circ R = \{\langle 1, 0.8\rangle, \langle 2, 1.0\rangle, \langle 3, 0.8\rangle\}\ .$$

Example 58. Let C be a fuzzy set in the universe of discourse $X = [0,1]$ and let R be a binary fuzzy relation in X. Assume that $C(x) = x$ and $R(x,y) = 1 - |x - y|$. Using the definition of sup-min composition we get $(C \circ R)(y) = \sup_{x \in [0,1]} \min[x, 1 - |x-y|] = \frac{1+y}{2}$ for all $y \in [0,1]$. See Fig (8.3).

8.3.2. $\sup$–t composition

As we saw in 6 on page 77, while $\min$ is the standard fuzzy operator for conjunction, often specific applications call for the use of t–norms, such as the product. It is very common to replace the $\min$operator in composition with a t–norms to produce $\sup$–t composition.

Definition 38 (sup–t composition)**.** Let $R \in \mathcal{F}\,(X \times Y)$ and $S \in \mathcal{F}(Y \times Z)$. The $\sup$-t composition of R and S, denoted by $R \circ S$ is defined as

$$(R \circ_{\mathrm{t}} S)(x,z) = \sup_{y \in Y} \mathrm{t}\,[R(x,y), S(y,z)]$$

Definition 39. The sup-t composition of a fuzzy set $C \in \mathcal{F}(X)$ and a fuzzy relation $R \in \mathcal{F}\,(X \times Y)$ is defined as

$$(C \circ_{\mathrm{t}} R)\,(y) = \sup_{x \in X} \mathrm{t}\,[C(x), R(x,y)]$$

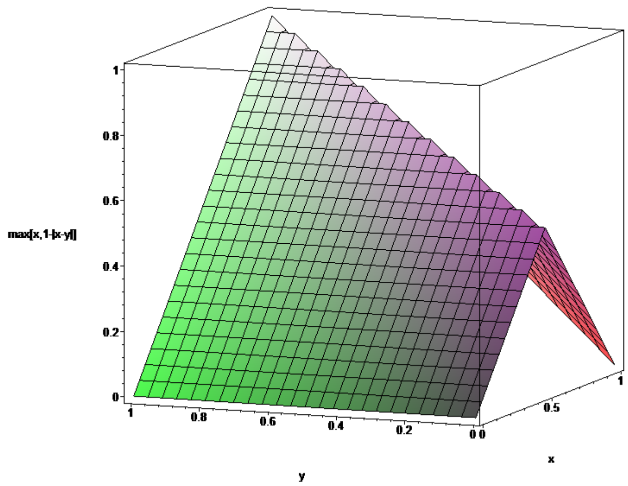


Figure 8.3.: Sup-min composition of fuzzy set and fuzzy relation.

for all y $\in Y$.

Definition 40. The sup-t composition of a fuzzy relation $R \in \mathcal{F}\ (X \times Y)$ and a fuzzy set $D \in \mathcal{F}(Y)$ is defined as

$$(R \circ_{\mathsf{t}} D)\,(x) = \sup_{y \in Y} \mathsf{t}\,[R(x,y), D(y)]$$

for all y $\in Y$.

It is clear that $R \circ_{\mathsf{t}} S$ is a binary fuzzy relation in $X \times Z$.

Example 59. Let C be a fuzzy set in the universe of discourse $\{1,2,3\}$ and let R be a binary fuzzy relation on $\{1,2,3\}$. Assume that $C = \{\langle 1, 0.2\rangle\,, \langle 2, 1.0\rangle\,, \langle 3, 0.2\rangle\}$ and that R is giving in the following table

R	1	2	3
1	1	0.8	0.3
2	0.8	1	0.8
3	0.3	0.8	1

. (8.12)

Suppose that we use the product t–norm. Using the definition of $\sup -\mathrm{t}$ composition we get

$$C \circ R = \begin{bmatrix} 0.2 & 0.4 & 0.5 \end{bmatrix} \circ_{\mathsf{t}} \begin{bmatrix} 1 & 0.8 & 0.3 \\ 0.8 & 1 & 0.8 \\ 0.3 & 0.8 & 1 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.4 & 0.5 \end{bmatrix}$$

so

$$C \circ R = \{\langle 1, 0.32\rangle, \langle 2, 0.4\rangle, \langle 3, 0.5\rangle\}\ .$$

8.3.2.1. Transitivity

As is typical in fuzzy set theory, fuzzification introduces some new concepts and new words into the lexicon. For example, suppose the relation R on X between elements a and b is stronger than any two step connection from a to b through some c. By this we mean that that the membership grade in the relation for the pair $\langle a, b\rangle$ is greater than the membership grade of both $\langle a, c\rangle$ and $\langle c, b\rangle$ for any all $c \in X$. This property of a relation is called sup-min transitivity .

Definition 41 (sup-min transitivity)**.** If R, $R : X \times X \to [0, 1]$, is a fuzzy relation then it is sup-min transitive if, for all $\forall x, y \in X$,

$$R(x, y) \geq \sup_{z \in Z} \min\left[R(x, z), R(z, y)\right] \tag{8.13}$$

In the Chapter (6) this book explained that the standard min operator of Zadeh's original fuzzy set theory is sometimes replaced by an alternate intersection operator, a t-norm. If we use this substitution in the definition of sup-min transitivity we get $\sup$-t transitivity .

Definition 42 (sup-t transitivity)**.** If R, $R : X \times X \to [0, 1]$, is a fuzzy relation and t is a t-norm then the relation R is $\sup$-t transitive if

$$R(x, y) \geq \sup_{z \in Z} \mathsf{t}\left[R(x, z), R(z, y)\right] \tag{8.14}$$

for all $\forall x, y \in X$.

Example 60. Let C be a fuzzy set in the universe of discourse $\{1\ , 2\ , 3\}$and let R be a binary fuzzy relation on $\{1, 2, 3\}$. Assume that $C = \frac{0.2}{1} + \frac{0.4}{2} + \frac{0.5}{3}$ and R is giving in the following table

R	1	2	3
1	1	0.8	0.3
2	0.8	1	0.8
3	0.3	0.8	1

. (8.15)

Suppose that we use the product t–norm. Using the definition of $\sup - \mathrm{t}$ composition we get

$$C \circ R = \begin{bmatrix} 0.2 & 0.4 & 0.5 \end{bmatrix} \overset{\mathsf{t}}{\circ} \begin{bmatrix} 1 & 0.8 & 0.3 \\ 0.8 & 1 & 0.8 \\ 0.3 & 0.8 & 1 \end{bmatrix} = \begin{bmatrix} 0.32 & 0.4 & 0.5 \end{bmatrix}$$

so

$$C \circ R = \frac{0.32}{1} + \frac{0.4}{2} + \frac{0.5}{3}\ .$$

The terms $\max$-$\min$ transitivity and $\max$-t transitivity are often used when $\sup$-$\min$ transitivity and $\sup$-t transitivity are technically correct. .

8.3.3. Types of fuzzy relations

Definition 43 (fuzzy equivalence relation)**.** If a fuzzy relation $R : X \times X \to [0, 1]$, is reflexive, symmetric, and sup–min transitive then it is called a fuzzy equivalence relation, or a similarity relation Zadeh (1971).

Example 61. The table following presents a fuzzy equivalence relation S on $X = \{a, b, c, d\}$

S	a	b	c	d
a	1.0	0.8	0.7	1.0
b	0.8	1.0	0.7	0.8
c	0.7	0.7	1.0	0.7
d	1.0	0.8	0.7	1.0

The α-cut of a fuzzy equivalence relation R are crisp equivalence relations.

Example 62. The α-cut of S given in Example (61) at $\alpha = 0.75$ is an equivalence relation as illustrated in the following table

${}^{0.75}S$	a	b	c	d
a	1	1	0	1
b	1	1	0	1
c	0	0	1	0
d	1	1	0	1

(8.16)

Example 63. The α-cut of S given in Example (61) at $\alpha = 0.85$ is an equivalence relation as illustrated in the following table

${}^{0.85}S$	a	b	c	d
a	1	0	0	1
b	0	1	0	0
c	0	0	1	0
d	1	0	0	1

(8.17)

An equivalence relation always produces a partition. In the example above (Eq. (8.17)) ${}^{0.85}S$ has partition classes $\{a, d\}$, $\{b\}$, and $\{c\}$. The partition corresponding to ${}^{0.75}S$ (Eq. (8.16)) has classes $\{a, b, d\}$ and $\{c\}$.

Note that as α increased in the two examples above the partition got finer. It is not hard to see, or to prove, that the family of partitions of the alpha-cuts of a fuzzy equivalence relation forms a nested sequence of partitions as α increases from zero to one. This means each element of the $\alpha = 0.75$ partition $\{\{a, d\}, \{b\}, \{c\}\}$ is a subset of the $\alpha = 0.85$ partition $\{\{a, b, d\}, \{c\}\}$.

Definition 44 (fuzzy preorder relationship)**.** A fuzzy relation that is reflexive, and $\sup$-$\min$ transitive is called a fuzzy preorder relationship on X.

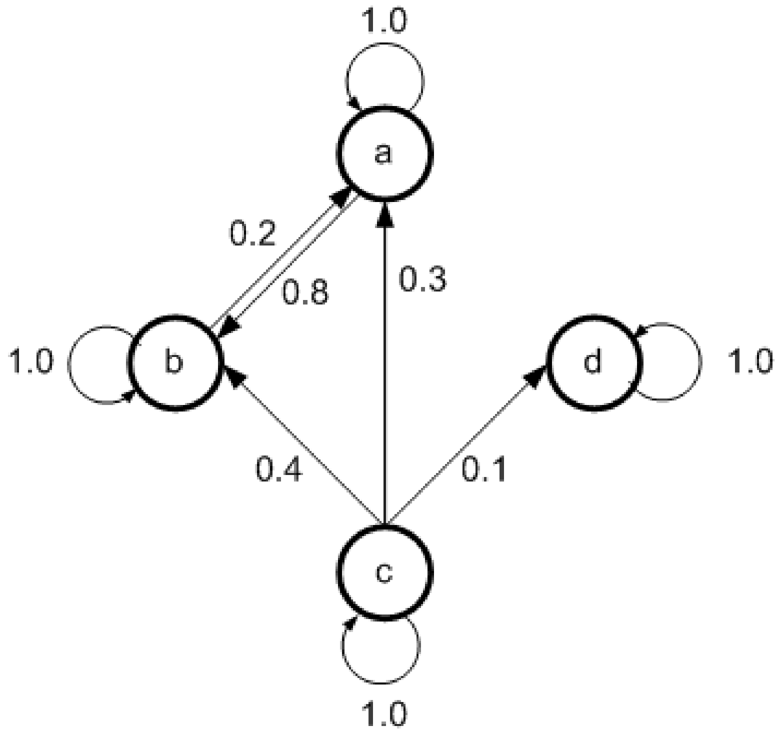


Figure 8.4.: Fuzzy order relation.

Example 64. The fuzzy relation R on $X = \{a, b, c\}$ specified in the following equation is a fuzzy preorder.

R	a	b	c
a	0.2	1.0	0.4
b	0.0	0.6	0.3
c	0.0	1.0	0.3

(8.18)

Definition 45 (fuzzy order relationship)**.** A fuzzy relation that is reflexive, antisymmetric, and sup-min transitive is called a fuzzy order relationship on X.

Example 65. Fig. (8.4) gives a graphical version of a fuzzy order relation on $X = \{a, b, c, d\}$.

Definition 46 (fuzzy partial order)**.** A fuzzy relation that is reflexive, perfectly antisymmetric, and sup-min transitive is called a perfect fuzzy order relationship on X or a fuzzy partial order relationship on X.

Definition 47 (fuzzy linear order)**.** A fuzzy order such that for all $x, y \in X$; $x \neq y$ implies that either $R(x, y) = 0$ or equivalently $R(y, x) = 0$ is called a fuzzy linear order on X (also called a total fuzzy order relation).

Definition 48 (compatibility relationship)**.** A fuzzy relation that is reflexive and symmetric is called a compatibility relationship on X.

8.4. Graph and fuzzy graph

Graphs can be used to represent relationships, both crisp and fuzzy. This chapter has already presented relationships graphically without a precise definition of a graph.

Definition 49 (graph)**.** A graph G is defined as an ordered pair: $G = (V, E)$ where
V : Set of **vertices**. A vertex is also called a node or element.
E : Set of **edges**. An edge is an unordered pair (x, y) of vertices in V.

It is important to note that (x, y) is an unordered pair. This means that $(x, y) = (y, x)$. When we consider ordered pairs the structure is called a directed graph or digraph.

Definition 50 (digraph)**.** A digraph G is defined as an ordered pair: $G = (V, E)$ where
V : Set of **vertices**. A vertex is also called a node or element.
E : Set of **edges**. An edge is an *ordered* pair $\langle x, y\rangle$ of vertices in V.

A digraph is a data structure that expresses a relationship since we can consider the edge set as a related pair. If $\langle a, b\rangle \in E$ then $a\ R\ b$ and $R \equiv E \subseteq V \times V$. In a regular graph the edge (a, b) is an unordered pair and corresponds to both $a\ R\ b$ and $b\ R\ a$ so that a graph expresses only symmetric relations.

When order is not allowed, we call a graph an undirected graph.

A path from x to y is a set of edges with continuous edges. If we set $a_0 = x$ and $a_n = y$ then a path is a set of edges (a_0, a_1), (a_1, a_2), (a_2, a_3), $\cdots$, (a_{n-1}, a_n) with $n \geq 1$ and $(a_i, a_{i+1}) \in E$ for all $0 \leq i \leq n - 1$. The length of path is a the number of edges in this path (n in the previous example). When there exists a path from node a to b in G, a and b are said to be connected. If all a, $b \in V$ in graph G are connected, the graph G is said to be a connected graph.

When sets X and Y (including the case where $X = Y$) are given along with a crisp relation R then we can define a directed graph $G_R(X \cup Y, E)$ where $\langle x, y\rangle \in E$ iff $x\ R\ y$.

Definition 51 (fuzzy graph)**.** A fuzzy graph G is defined as an ordered pair: $G = (V, E)$ where
V : Set of **vertices**. A vertex is also called a node or element.
E : Set of fuzzy **edges**. An edge is an element of the fuzzy set $E : X \times Y \to [0, 1]$.

If we are given a fuzzy relation $R \subseteq \mathcal{F}(X \times Y)$ then we can define the fuzzy graph $G_R(X \cup Y, E)$, where $E \equiv R$. A fuzzy graph is essentially a classical weighted graph, with the weights restricted to the unit interval.

A fuzzy graph is an expression of fuzzy relation and thus the fuzzy graph is frequently expressed as fuzzy matrix.

Example 66. Fig (8.5) shows an example of fuzzy graph representing the fuzzy relation matrix M_G given in Table (8.3).

8.5. Operations on fuzzy relations

Since fuzzy relations are simply fuzzy subsets of $X \times Y$ we can perform the familiar fuzzy set operations upon them. For example, we can take the complement of a fuzzy relation R. If we have two fuzzy relations R and S, both on $X \times Y$ then we can calculate their intersection or union.

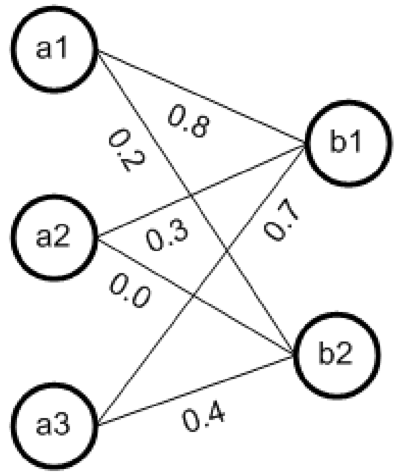


Figure 8.5.: The fuzzy relation M_G as a graph.

M_G	b_1	b_2
a_1	0.8	0.2
a_2	0.3	0.0
a_3	0.7	0.4

Table 8.3.: A fuzzy relation matrix M_G.

Definition 52. Let R be binary fuzzy relations on $X \times Y$. The complement of R is defined as

$$R^{\mathsf{c}}(x, y) = 1 - R(x, y)\,. \tag{8.19}$$

Definition 53. Let R and S be two binary fuzzy relations on $X \times Y$. The intersection of R and S is defined by

$$(R \cap S)(x, y) = min\,[R(x, y), S(x, y)]\;. \tag{8.20}$$

Definition 54. Let R and S be two binary fuzzy relations on $X \times Y$. The union of R and S is defined by

$$(R \cup S)(x, y) = max\,[R(x, y), S(x, y)]\,. \tag{8.21}$$

Example 67. Let us define two binary relations

$R =$ "$\mathsf{x\ is\ considerable\ larger\ than\ y}$" $=$ (8.22)

R	y_1	y_2	y_3	$y4$
x_1	0.8	0.1	0.1	0.7
x_2	0	0.8	0	0
x_3	0.9	1	0.7	0.8

and

$S =$ "x is very close to y" $=$

S	y_1	y_2	y_3	y_4
x_1	0.4	0	0.9	0.6
x_2	0.9	0.4	0.5	0.7
x_3	0.3	0	0.8	0.5

(8.23)

The intersection of R and S means that "x is considerable larger than y" and "x is very close to y".

$(R \cap S)(x, y) =$

$R \cap S$	y_1	y_2	y_3	y_4
x_1	0.4	0	0.1	0.6
x_2	0	0.4	0	0
x_3	0.3	0	0.7	0.5

The union of R and S means that "x is considerable larger than y "or "x is very close to y".

$(R \cup S)(x, y) =$

	y_1	y_2	y_3	y_4
x_1	0.8	0	0.9	0.7
x_2	0.9	0.8	0.5	0.7
x_3	0.9	1	0.8	0.8

8.5.1. Shadows

In 2.7 on page 20 we examined what happens when you take a set of ordered pairs and look at only the elements in the first or second dimension. These were called projections. We will now do something similar for fuzzy sets defned on ordered pairs. The reults are called shadows.

Definition 55. Let R be a binary fuzzy relation on $X \times Y$. The shadow of R on X is a fuzzy set defined upon X with a membership function given by

$$shad_1 R(x) = \sup_{y \in Y} [R(x, y\ )] \tag{8.24}$$

and the shadow of R on Y is dis a fuzzy set defined upon Y with a membership function given by

$$shad_2 R(y) = \sup_{x\ \in X} [R(x, y\ )] \tag{8.25}$$

If X and Y are finite sets then $\sup$, for supremum, can be replaces with the more familiar $\max$.

Example 68. Consider the relation R = "x is considerable larger than y" whose numerical values are given above by Eq. (8.22), then the shadow on X means that x_1 is assigned the highest membership degree from the grades of the tuples $\langle x_1, y_1 \rangle$, $\langle x_1, y_2 \rangle$, $\langle x_1, y_3 \rangle$, and $\langle x_1, y4 \rangle$ in R, i.e. $shad_1 R(x_1) = 1$, which is the maximum of the first row. x_2 is assigned the highest membership degree from the tuples $\langle x_2, y_1 \rangle$, $\langle x_2, y_2 \rangle$, $\langle x_2, y_3 \rangle$, and $\langle x_2, y4 \rangle$, i.e. $shad_1 R(x_2) = 0.8$, which is the maximum of the second row. . x 3 is assigned the highest membership degree from the tuples $\langle x_3, y_1 \rangle$, $\langle x_3, y_2 \rangle$, $\langle x_3, y_3 \rangle$, and $\langle x_3, y4 \rangle$, i.e. $shad_1 R(x_3) = 1$, which is the maximum of the third row.

8.5.2. Transitive closure

Let R be a binary relation on a set X. It is always possible to construct a relation that is $\sup$-$\min$ transitive, based upon R, by forming the transitive closure of R. To form the transitive closure of R, set $C = R$ and then repeatedly form $D = (C \circ C) \cup C$ until $D = C$, which it eventually must. This relation, C, is then the transitive closure of R.

Definition 56. Let R be a binary relation on a finite set X. Set $C = R$ and the repeatedly form $D = (C \circ C) \cup C$ until $C = D$. The transitive closure of R is then C.

Example 69. Let R be the fuzzy relation:

$$R = \begin{bmatrix} 0.70 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.0 \\ 0.00 & 0.40 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.80 & 0.00 \end{bmatrix}$$

Set $C = R$ and and form $D = (C \circ C) \cup C$:

$$D = \begin{bmatrix} 0.70 & 0.50 & 0.00 & 0.50 \\ 0.00 & 0.00 & 0.80 & 1.0 \\ 0.00 & 0.40 & 0.00 & 0.40 \\ 0.00 & 0.40 & 0.80 & 0.00 \end{bmatrix}$$

Since $D \neq C$ replace C with D and calculate the new $D = (C \circ C) \cup C$:

$$D = \begin{bmatrix} 0.70 & 0.50 & 0.50 & 0.50 \\ 0.00 & 0.40 & 0.80 & 1.0 \\ 0.00 & 0.40 & 0.40 & 0.40 \\ 0.00 & 0.40 & 0.80 & 0.40 \end{bmatrix}$$

Again $D \neq C$ so we replace C with D and calculate the next $D = (C \circ C) \cup C$:

$$D = \begin{bmatrix} 0.70 & 0.50 & 0.50 & 0.50 \\ 0.00 & 0.40 & 0.80 & 1.0 \\ 0.00 & 0.40 & 0.40 & 0.40 \\ 0.00 & 0.40 & 0.80 & 0.40 \end{bmatrix}$$

This is the same result as before so D is the transitive closure of R.

The same process can be applied using any $\sup$-t composition to form the $\sup$-t transitive closure.

8.5.3. Distances and fuzzy relations

Distances 2.6 on page 18 can be used to construct fuzzy relations. The notion here is that as things are closer, the are more related. The trick is the distance goes the wrong way, and is potentially infinite. When the distance between A and B is zero the objects are identical, which is very similar, so we want the relationship grade to be one. As the distance gets bigger, and A and B are farther apart, we want

Algorithm 8.1 Transitive Closure

Require: A set of data points X of size n.
Require: A relation R on X
Set $D = R$.
repeat
  $C \Leftarrow D$
  $D \Leftarrow (C \circ C) \cup C$
until $C = D$
{It can be proved that the algorithm will always terminate}
Return $R^* = D$, the transitive closure of R

the relationship grade to decrease towards zero. The following makes this intuitive notion precise.

For fuzzy set theory, the important thing about distances is, given two points x_1 and x_2, both elements of X, the distance, $dist(x_1, x_2)$, is a non-negative real number. Suppose $f(z)$ is a monotonic (usually decreasing) function that maps the non-negative reals into the unit interval, i.e., $Z = \mathbb{R}^+$ and for $z \in Z$ we have that $f(z) \in [0, 1]$. There is now an instant conversion between any distance measure between points in X and a fuzzy relation on X. The membership degree of this relation R is given by the formula

$$R(x_1, x_2) = f\left(dist(x_1, x_2)\right).$$

Example 70. Let $Z = \mathbb{R}^+$ and $X = \mathbb{R}^2$. Define $f(z) = 1 - e^{-z}$ for $z \in Z$ and use the Euclidean distance $dist(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle)$ for $\langle x, y \rangle \in \mathbb{R}^2$. Then the fuzzy relation on X induced by f and $dist$ is given by

$$R(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) = 1 - e^{-\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

When $\langle x_1, y_1 \rangle = \langle 1, 5 \rangle$ and $\langle x_2, y_2 \rangle = \langle 4, 1 \rangle$ then $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = 5$ and

$$\begin{aligned} R(\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle) &= 1 - e^{-5} \\ &= 0.993\,. \end{aligned}$$

8.6. Notes

The basic ideas of fuzzy relations and the concepts of similarity and fuzzy ordering were introduced by Zadeh (1971). Binary fuzzy relations were further investigated by Rosenfeld (1975), Yeh and Bang (1975), Yager (1981), and Ovchinnikov (1984). There is an excellent chapter on fuzzy relations in Kaufmann (1975).

The concepts of projection, cylindric extension, and cylindric closure appeared in their crisp version in Ashby (1964) and in their fuzzy versions in Zadeh (1975b) and Zadeh (1975a).

8.7. Homework

Given the universes
$$\begin{aligned} X &= \{0,1,2,3,4,5,6,7,8,9,10\} \\ Y &= \{1,2,3\} \\ Z &= \{a,b,c\} \\ U &= \{\alpha,\beta,\gamma\} \\ V &= \{0,1,2,3,4,5,6,7,8,9,10\} \end{aligned}$$
and the fuzzy sets and fuzzy relations

$$\mathsf{A}(x) = \frac{x}{10}$$
$$\mathsf{B}(x) = \frac{|x-5|}{5}$$
$$\mathsf{C}(x) = \begin{cases} \frac{9-|2x-9|}{8} & 1 \le x \le 8 \\ 0 & otherwise \end{cases}$$
$$\mathsf{D}(x) = 0.8$$
$$\mathsf{E}(x) = \{\langle 1, 0.2\rangle, \langle 2, 0.6\rangle, \langle 3, 0.4\rangle\}$$
$$\mathsf{F}(y) = \{\langle a, 0.3\rangle, \langle b, 0.7\rangle, \langle c, 0.9\rangle\}$$
$$\mathsf{G}(z) = \{\langle \alpha, 0.5\rangle, \langle \beta, 0.8\rangle, \langle \gamma, 0.35\rangle\}$$

R	a	b	c
1	1.0	0.5	0.1
2	0.4	1.0	0.2
3	0.6	0.5	1.0

S	α	β	γ
a	0.9	0.5	0.1
b	0.4	0.7	0.2
c	0.6	0.5	0.9

T	α	β	γ
α	0.8	0.4	0.4
β	0.4	0.7	0.7
γ	0.6	0.5	0.9

E	a	b	c	d	e	f
a	1.0	0.8	0.0	0.0	0.0	0.0
b	0.8	1.0	0.0	0.0	0.0	0.0
c	0.0	0.0	1.0	1.0	0.8	0.0
d	0.0	0.0	1.0	1.0	0.8	0.7
e	0.0	0.0	0.8	0.8	1.0	0.7
f	0.0	0.0	0.0	0.7	0.7	1.0

F	Γ	Δ	Θ	Φ
Γ	1.0	0.2	0.2	0.6
Δ	0.2	1.0	0.4	0.2
Θ	0.2	0.4	1.0	0.2
Φ	0.6	0.2	0.2	1.0

Given the above information, try to answer the following questions.

1. What is $A \times B$?
2. What is $B \times A$?
3. What is $A \times C$?
4. What is $C \times A$?

5. What is $C \times B$?
6. What is $B \times C$?
7. What is $D \circ R$?
8. What is $D \circ S$?
9. What is $E \circ R$?
10. What is $E \circ S$?
11. What is $F \circ R$?
12. What is $F \circ S$?
13. What is $S \circ T$?
14. What is $T \circ T$?
15. What is $S \circ T \circ T$?
16. What is $T \circ T \circ T$?
17. What is $E \circ E$?
18. What is $E \circ E \circ E$?
19. What is the projection of the fuzzy set R into X?
20. What is the projection of the fuzzy set R into Y ?
21. What is the cross product of the projections of R into X and Y? How does this result compare with the original fuzzy relation R?
22. What is the projection of the fuzzy set S into Y?
23. What is the projection of the fuzzy set S into Z ?
24. What is the cross product of the projections of S into Y and X? How does this result compare with the original fuzzy relation S?
25. What is the shadow of the fuzzy set T into Z in the first dimension?
26. What is the shadow of the fuzzy set T into Z in the second dimension?
27. What is the cross product of the shadows of T into Z in the first and second dimensions? How does this result compare with the original fuzzy relation T?
28. What properties (symmetric, reflexive, etc., see Table 8.2) does T possess?
29. What properties (symmetric, reflexive, etc., see Table 8.2) does E possess?
30. Is E a similarity relation? What is its image set? List the α–cuts of E for each value α in the image set.

31. Let H be the set of your High School friends. Let W be the relation *walking distance*. Decide how you will define the relation g is within *walking distance* to h. Draw a graph the result relation, as in Fig. (8.4). What kind of relation is W.

32. Let H be the set of your High School friends. Let Q be the relation is *friends with*. Decide how you will define the relation g is *friends with* h. Draw a graph the result relation, as in Fig. (8.4). What kind of relation is W.

9. Fuzzification

9.1. Introduction

A cook is only as good as his ingredients. A mechanic is only as good as his tools. It is true that a better chef will cook a better dish with the same ingredients and a better mechanic will do a better repair job with the same set of tools. But it is hard to cook anything good from rotten vegetables, and it is hard to fix a flat tire without a lug-wrench, a jack and a spare tire.

While fuzzy sets are used to represent indeterminacy and are not designed to be a precise representation, it is also true that a poor representation will create a system that is useless for any constructive purpose. A system that represents young as a fuzzy set centered on the ages 90–100 would not be appropriate even in a geriatric facility.

An engineer who is using fuzzy sets is seeking an optimal controller. An computer scientist who is using fuzzy sets for data mining is looking for a superior sieve. An analyst who is using fuzzy sets for clustering is looking for a valuable association.

Definition 57 (fuzzification)**.** The construction and design of fuzzy set membership functions.

The question then is how do we construct good fuzzy sets from data and how do we change the fuzzy sets adoptively. The methods fall into three broad categories; manual, automatic, and adaptive. Manual methods primarily deal with evidence obtained from human responses. Automatic methods are primarily used for processing data sets to determine appropriate fuzzy set representation. Adaptive methods search for an optimal system design. In this chapter we will examine some of the most common methods in use for the design and construction of fuzzy sets.

9.2. What is a fuzzy set?

There are three major ways of looking at a fuzzy set for the purpose of construction: vertical, horizontal, and random set.

9.2.1. Vertical

The vertical view of a fuzzy set is the membership function view. For each $x \in X$ the membership function $\mu_A(x)$ gives the degree of compatibility of the element x with the category or linguistic term A presently under discussion. It has always been the assumption that the greater the membership value, $\mu_A(x)$, the greater the compatibility of x with the concept represented by A. Therefore to construct $\mu_A(x)$ vertically from information about x, it is necessary that the data allows one to order the compatibility of x with the concept A.

Class Worksheet Card ______

1. Circle the number in the scale 0-10 that represents how much you agree with the statement.

I.1	A five year old human is young.	0 1 2 3 4 5 6 7 8 9 10
I.2	A five year old human is very young.	0 1 2 3 4 5 6 7 8 9 10
I.3	A six year old human is young.	0 1 2 3 4 5 6 7 8 9 10
I.4	A four year old human is young.	0 1 2 3 4 5 6 7 8 9 10
I.5	A four year old human is very young.	0 1 2 3 4 5 6 7 8 9 10
I.6	A six year old human is very young.	0 1 2 3 4 5 6 7 8 9 10
I.7	A seven year old human is young.	0 1 2 3 4 5 6 7 8 9 10
I.8	A three year old human is young.	0 1 2 3 4 5 6 7 8 9 10

II Answer the following questions with numerical values between 1 and 100 years.

II.1	What range of ages do you consider young?	______ to ______
II.2	What range of ages do you consider old?	______ to ______
II.3	At what height do you consider a woman short?	______
II.4	What range of ages do you consider middle aged?	______ to ______
II.5	What age do you think of when someone says infant?	______
II.6	What age do you think of when someone says ancient?	______
II.7	What range of ages do you consider the prime of life?	______ to ______
II.8	What range of ages do you consider very young?	______ to ______
II.9	What range of ages do you consider not old?	______ to ______
II.10	What age do you think of when someone says geriatric?	______
II.11	What age do you think of when someone says child?	______
II.12	What range of ages comes to your mind when you here someone say "not very young?"	______ to ______

III For the following questions, rate your agreement with a Likert Scale:

Strongly disagree	$\longleftrightarrow$	1
Disagree	$\longleftrightarrow$	2
Neutral	$\longleftrightarrow$	3
Agree	$\longleftrightarrow$	4
Strongly agree	$\longleftrightarrow$	5

III.1	I would pay $9 for a good steak dinner.	______
III.2	I would pay $8-$10 for a good steak dinner.	______
III.3	I would pay $7-$11 for a good steak dinner.	______
III.4	I would pay $6-$12 for a good steak dinner.	______
III.5	I would pay $5-$13 for a good steak dinner.	______

9.2.1.1. The standard vertical transformations

Since the membership function of a fuzzy set always takes values in $[0,1]$ the standard vertical transformation of any function $f : X \to \mathbb{R}$ (or data set) into a fuzzy set F with membership μ_F is

$$\mu_F(x) = \frac{f(x) - f_{\min}}{f_{\max} - f_{\min}}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum values obtained by the function (or data) over the domain set X. If the function maps to the nonnegative reals, $f : X \to \mathbb{R}^+$, or the data are all positive, then it is sometimes assumed that $f_{\min} = 0$ and the formula

$$\mu_F(x) = \frac{f(x)}{f_{\max}}$$

is used.

9.2.2. Horizontal

The horizontal view of a fuzzy set is the α–cut. This is an interval associated with each $\alpha \in [0,1]$. If we think of α as a confidence factor or surety level, then we can interpret the α–cut as the x values that we are at least α sure are compatible with the concept or label A. The decomposition theorem 5.39 ensures that the fuzzy set membership function can be constructed from the parametric family of its α–cuts. Therefore, if one can determine the α–cuts then one can determine the fuzzy set. When α is 1 we need to determine the smallest and largest values of x that are completely compatible with the concept or category A. When α is 0.5 we need to determine the smallest and largest values of x that are half-way compatible with or 50% surely possess the concept or category A.

9.2.2.1. The standard horizontal transformations

When we determine enough α-cuts, ${}^{\alpha}A$, we can produce a model of the membership function $\mu_A(x)$ using the formula

$$\mu_A(x) = \sup_{\alpha \in [0,1]} \{\alpha \mid x \in {}^{\alpha}A\}$$

9.2.3. Random set

Suppose that X is discrete. Even if X is not discrete, it can almost always be represented by a discrete sampling. DVDs and CDs show that continuous phenomena, like acting and music can be digitized. Digitization turns continuous phenomena into discrete phenomena.

Example 71. Let us take a simple triangular fuzzy number one, $Tr[0,1,2]$. If we sample this function every $\frac{1}{4}$ of a unit form 0 to 2 we get the discrete fuzzy number,

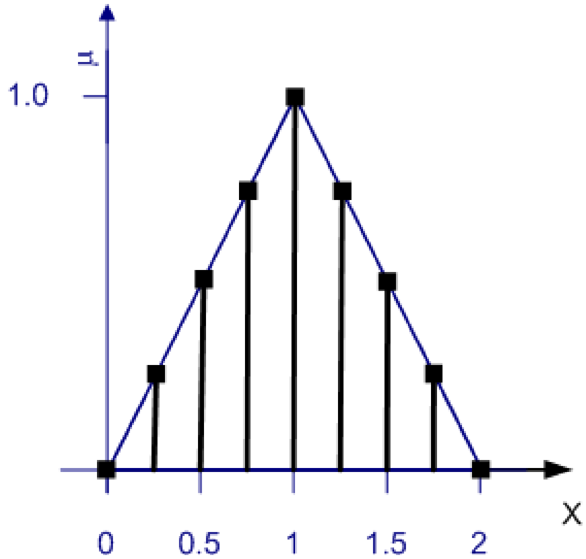


Figure 9.1.: Triangular fuzzy number Tr and its discrete version.

using fraction notation,

$$Tr' = \frac{0.0}{0.0} + \frac{0.25}{0.25} + \frac{0.5}{0.5} + \frac{0.75}{0.75} + \frac{1.0}{1.0} + \frac{0.75}{1.25} + \frac{0.5}{1.5} + \frac{0.25}{1.75} + \frac{0.0}{2.0} \tag{9.1}$$

as seen in Fig. (9.1).

Let A be a fuzzy set on discrete X. Let $I(A)'$ be the image set of A minus the set $\{0\}$ and with the membership grades sorted in decreasing order:

$$\Lambda_{(A)} = [\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_n]$$

where $\alpha_1 > \alpha_2 > \alpha_3 > \cdots > \alpha_n > 0$ and set $\alpha_{n+1} = 0$ where $n = \left|\Lambda_{(A)}\right|$. Define $m_i = \alpha_i - \alpha_{i+1}$ for $i \in \mathbb{N}_n$. Then the random set representation of A is

$$\rho_A = \{\langle {}^{\alpha_i}A, m_i\rangle \mid \alpha_i \in \Lambda'_A\} \ .$$

We can reconstruct the membership function of A from its random set representation:

$$\mu_A(x) = \sum_{x \in {}^{\alpha_i}A} m_i$$

where $\langle {}^{\alpha_i}A, m_i\rangle \in \rho_A$. There is also a connection between the random set interpretation and the possibility theory interpretation of fuzzy set theory.

Example 72. For the discrete fuzzy set Tr' of Eq. (9.1) the image set is in decreasing order with zero omitted is

$$\Lambda_{(Tr')} = [1.0, 0.75, 0.5, 0.25] \tag{9.2}$$

and the random set representation is

$$\rho_{Tr'} = \left\{ \begin{array}{c} \langle \{1.0\}, 0.25 \rangle, \\ \langle \{0.75, 1.0, 1.25\}, 0.25 \rangle, \\ \langle \{0.5, 0.75, 1.0, 1.25, 0.75, 1.5\}, 0.25 \rangle, \\ \langle \{0.25, 0.5, 0.75, 1.0, 1.25, 0.75, 1.5, 1.75\}, 0.25 \rangle \end{array} \right\} . \tag{9.3}$$

We can recover the membership grades by adding the m values for each set in $\rho_{Tr'}$ that contains x.

$$Tr'(0.5) = 0.25 + 0.25 \,. \tag{9.4}$$

9.2.4. Parametric fuzzy sets

There is one other common method of constructing fuzzy sets. If experience has shown that triangular, or trapezoidal, or some other basic shape of fuzzy set seems to work in a particular application, then mathematical techniques can be used to search for the optimal parameters of the basic fuzzy set shape. These techniques are similar to finding a trend line or regression line in statistics.

For example, if triangular fuzzy sets will work in a particular application, then they always have the form $Tr[a, m, b]$. Thus Tr is a function that has three parameters: a, m, and b. There are many mathematical search techniques, such as least square error fitting, that may be able to find the appropriate values of a, m, and b given an appropriate data set.

Example 73. This is essential the reverse of the digitization process illustrated in Fig. 9.1. The idea here is to take the data pairs presented in Zadeh fraction notation in Eq. (9.1) and recover the triangular fuzzy set function

$$Tr[0, 1, 2] = \left\{ \begin{array}{cc} x & 0 < x < 1 \\ 2 - x & 1 \leq x < 2 \end{array} \right. . \tag{9.5}$$

9.3. Manual methods

Various statistical techniques are often used in the determination of fuzzy set membership functions. In Watanabe (1993) the author asserts that these statistical techniques fall into two broad categories: the use of frequencies and direct estimation. The first methodology, the frequency method, obtains the membership function by measuring the percentage of people in a test group (typically experts in the particular domain under consideration) who answer affirmatively to a question about whether an object belongs to a particular set. The second methodology, direct estimation, derives its values from a sliding scale, it elicits a responses from experts that grade the compatibility of the object and the set. Experiments conducted by Watanabe came to the conclusion that direct estimation methods are superior to the frequency method. In Turksen (1991) the author examined four different approaches (that include Watanebe's) to the acquisition of membership functions: *direct rating*, *polling*, *set valued statistics* and *reverse rating*. Turksen uses a system that is very similar to the linguistic system $\langle x, T(x), U, G, M \rangle$ introduced in Chapter (16) His notation includes:-

1. The set of elements $\theta \in \Theta$, the domain, (U in Zadeh's linguistic system);
2. V, a linguistic variable which is a label for $\theta \in \Theta$ (x in Zadeh's linguistic system);
3. A, a linguistic term of a linguistic variable ($T(x)$ is the set of terms in Zadeh's linguistic system);
4. A measurable numerical assignment interval $X \in [-\infty, \infty]$;
5. $\mu_A(\theta)$, the membership value representing the degree of membership of θ to the set of elements determined by linguistic term A, ($M_A(u)$ in Zadeh's linguistic system).

For example when considering the set of men we have a 'man' , θ, in the set 'men', Θ, whose height, V, describes how 'tall', A, he is. For example the height might be in the range $[0, 4]$ meters, X, and how tall Michael Jordan might be is the value of the membership function $\mu_{\mathsf{tall}}(MJ)$. *Direct rating* presents randomly selected $\theta \in \Theta$, with values $V(\theta) \in X$ to subjects who answer the question "How A is $\theta\,(V(\theta))$?" . In other words the question put to the expert is "How tall is Michael Jordan?" (note that 'tall' is different from 'height') and they respond by using a simple indicator on a sliding scale. Then, using the experts' opinion of the range of heights, a simple calculation reveals $\mu_{\mathsf{tall}}(MJ)$. This experiment is repeated for other men and the same candidate, Michael Jordan, repeatedly to reduce error. *Polling* asks a different question "Do you agree that θ is A?" expecting a yes or no response. The ratio of yes responses to total responses is used to arrive at a proportion that is then used to help generate the membership function. So, in our example, many respondents would be asked "Do you agree that Michael Jordan is tall?". *Set valued statistics* rely on the idea of combining ordinary sets using a frequentist approach based on observation. This method asks questions such as "Is the range 6 feet to 7 feet correctly termed tall?" Frequencies gained from asking this question repeatedly are then amalgamated to produce a fuzzy set membership function. *Reverse rating* takes a different approach by asking an expert to answer the following question "Identify $\theta(V(\theta))$ that has the y-th degree of membership in fuzzy set A." So we might ask an expert "Identify a man whose height indicates that he possesses the degree 0.5 of membership in the fuzzy set tall. Again, providing the expert has an understanding of the upper and lower limits this allows for a ready representation of μ_{tall}. Other authors (e.g. Kempton (1984)) employ various knowledge acquisition interviewing techniques for acquiring membership functions.

All the 'manual' approaches suffer from the same deficiency; they rely on the very subjective interpretation of words, the foibles of human experts, and generally, all the knowledge acquisition problems that are well documented with knowledge based systems.

9.4. One Expert

If there is one expert available who can manage a system and the goal is to automate his skill before he retires then direct questioning is the best available method. However the elicitation of knowledge from an expert is a very difficult and specialized skill. An expert may be cooperative or antagonistic, he may not consciously have numerical values associated with the rules he uses to control the process. He may not even have encapsulated his knowledge as rules, since he may not have trained a replacement. Furthermore, some rules may be used so seldom that only the dire circumstances of their necessity can bring them forth from long term memory.

This is very similar to the problems posed by constructing inference engine expert systems or knowledge based systems (KBS). In fact this problem is similar to the common computer science dilemma of requirements specification. The elicitation of knowledge from the problem domain and its experts, who speak their own language, is a difficulty common to all software engineering. As such, the literature on these subjects does give some guidelines.

First of all patience is required. The expert may be trying to be helpful, even if his answers are not taking the form that you like. It may be necessary to try many different approaches, and many different questions to infer the shape of a fuzzy set. Practice has shown that triangular and trapezoidal fuzzy sets work very well in many applications so that the elicitation of critical values with questions like:

> What rate of change in velocity causes you to start paying close attention?
> Ans: a.
> What rate of change in velocity causes you to start action B_{ji}?
> Ans: b.
> What rate of change in velocity causes you to start an action other than B_j?
> Ans: c.
> What rate of change in velocity generally indicates the process is out of control?
> Ans: d.

It is important *not* to put the questions in direct order, and possibly *not* to ask them all at the same time. Cognitive psychology has many things to say about the best methodologies of questioning people. However, if the answers to these questions form an order $a \leq b \leq c \leq d$ then $A_i = Tp\,[a, b, c, d]$ is a good fuzzy number to represent the critical velocity. Note that experts give us both knowledge of the fuzzy set memberships as well as (possibly) knowledge of the inference rules for the construction of an Approximate Reasoning system.

In other cases it may be difficult for the expert to give precise numerical values to such questions, however range values are almost as good, and sometimes all that is really needed is to find the core of the fuzzy numbers even if this is just a single value. A group of statement such as:

> "When the temperature is about $110°$ then lower the boiler feed by about two liters of foo."
> "When the temperature is about $125°$ then lower the boiler feed by about three liters of foo."
> "When the temperature is about $140°$ then lower the boiler feed by about six liters of foo."

gives the critical central values $a_1 = 110$, $a_2 = 125$, and $a_3 = 140$ for triangular fuzzy numbers $A_i = Tr\,[a_i - s_i, a_i, a_i + s_i]$ and $b_1 = 2$, $b_2 = 3$, and $b_3 = 6$ for triangular fuzzy numbers $B_j = Tr\,[b_j - s_j, b_j, b_j + s_j]$. The spread values of s_i and s_j, $i, j \in \mathbb{N}_3$, can be determined graphically using the granularity heuristic that the fuzzy sets A_i should cover the critical range of the input space and have a 10—20% overlap with the neighbors on each side. Thus if the critical range of temperatures is $100°$—150 then the three antecedent fuzzy sets

$$A_1 = Tp\,[100, 110, 120] \tag{9.6}$$

$$A_2 = Tr\,[115, 125, 135] \tag{9.7}$$

and

$$A_1 = Tp\,[130, 150, 150] \tag{9.8}$$

partition the input interval appropriately. The fuzzy sets B_1, B_2, and B_3, can be similarly constructed once the range of its domain set is determined.

9.5. Many Experts

The problem of determining fuzzy set membership grades when many experts are available has both positive and negative aspects. On the plus side, a weighted average of values obtained from the group of experts certainly contains a greater amount of information than that derived from a single person. However, the determination of appropriate weights and the melding of potentially conflicting evidence presents difficulties.

Suppose that there are n experts available. Assume that the experts evidence comes as *polling*. We are asking the question “Does θ belong to A?” or, equivalently, “Does x belong to A?” where θ has value x. That is, Michael Jordan has height $6'\,7''$ so one can ask, “Does Michael Jordan belong to tall?” or, equivalently, “Does $6'\,7''$ belong to $tall$?” Then each expert $i (i \in \mathbb{N}_n)$, gives a value $a_i(x)$ for each $x \in X$ that is either 0 if the expert does not believe the statement is true, or 1 if the expert does believe the statement is true. We can then use the simple approximation

$$\mu_A(x) = \frac{\sum_{i=1}^{n} a_i(x)}{n} \tag{9.9}$$

for the set membership grade. If on the other hand the values $\tilde{a}_i(x)$ are derived from *direct polling* then they may not be values in the unit interval. However the linear transformation

$$a_i(x) = \frac{\tilde{a}_i(x) - a_{\min}}{a_{\max} - a_{\min}} \tag{9.10}$$

where $a_{\max}$ and $a_{\min}$ represent the maximum and minimum values that the scale allows, will transform the data so that Eq. (9.9) is applicable.

In the absence of additional information we must assume that all experts are equally qualified. Of course there may be reason to trust more experienced experts more than their less seasoned colleagues.. The uniform weight of $c_i = \frac{1}{n}$ represents equal confidence in all experts. We can then express the Eq. (9.9) in the form

$$\mu_A(x) = \sum_{i=1}^{n} c_i \cdot a_i(x) \tag{9.11}$$

However, if years of experience or breadth of knowledge is incorporated in a weighing system other than uniform then the values of c_i, $i \in \mathbb{N}_n$, should be nonnegative and add to one. Here again the values of $a_i(x)$ may come from polling or from direct estimation.

9.5.1. Indirect methods

Both the method for one expert and the method for multiple experts can be extrapolated to the case where the data is specified as intervals. Turksen calls this *set valued statistics*. As mentioned previously, an expert will probably refuse to give information he believes is misleading. This may cause him to make statements such as;

> "When the temperature is between $120°$ and $130°$ then lower the boiler feed rate by about three liters of *foo*."

We can gather data about the intervals as easily as we gathered data about exact values. For example, the question "Is the range 6 feet to 7 feet correctly termed tall?" can be asked of each expert and the number of agreements for each interval calculated. Assume that m intervals are used in the questions or presented by the experts. Relative frequencies can then be calculated so that each interval gets a weight l_i equal to the number of agreements that interval L_i is correctly termed A divided by the number of agreements to all intervals. Finally the fuzzy set membership grade is calculated as

$$\mu_A(x) = \frac{\sum_{x \in L_i} l_i}{\sum_{i=1}^{m} l_i}\,. \tag{9.12}$$

This methodology has a strong connection to random set theory.

Another indirect method is based on the ability to compare objects. If there are n different values that we are going to use to determine the fuzzy set membership function, then it may be easier to get answers to questions of the type, "How much more A is x_i then x_j?" For example, "How much more $Taller$ is *Reggie Miller* then *Cheryl Miller*?" If we can determine a preference matrix $\mathbf{P} = [p_{ij}]_{nn}$ that is consistent (consistency means $p_{ij} = p_{ik} \cdot p_{kj}$) then we may be able to determine the individual membership grades $m_i = \mu_A(x_i)$ by assuming that $p_{ij} = \frac{m_i}{m_j}$. Under this assumption

$$\sum_{j=1}^{n} p_{ij} m_j = \sum_{j=1}^{n} m_i = n m_i \quad . \tag{9.13}$$

or in matrix form

$$\mathbf{Pm} = n\mathbf{m} \tag{9.14}$$

where the vector $\mathbf{m} = [m_i]_n^T$ are the membership grades to be determined. Equation (9.14) can be written

$$(\mathbf{P} - n\mathbf{I})\,\mathbf{m} = 0 \tag{9.15}$$

where $\mathbf{I}$ is the identity matrix. This equation has a solution if and only if $\mathbf{m}$ is an eigenvector of the matrix $\mathbf{P} - n\mathbf{I}$ and n is an eigenvalue. A final assumption is necessary to complete the derivation. If $\mathbf{m}$ is a solution to Eq. (9.15) than so is any scalar multiple. Hence a unique final solution comes only if we assume that $\max m_i = 1$ or some smaller positive value. Another method is to assume that $\sum m_i = 1$. In this case

$$\sum_{i=1}^{n} p_{ij} = \sum_{i=1}^{n} \frac{m_i}{m_j} = \frac{1}{m_j} \sum_{i=1}^{n} m_i = \frac{1}{m_j} \tag{9.16}$$

where

$$m_j = \frac{1}{\sum_{i=1}^{n} p_{ij}} \,. \tag{9.17}$$

If the data is slightly inconsistent then it may still be possible to find a solution to the linear equation

$$\mathbf{Pm} = \lambda \mathbf{m} \tag{9.18}$$

for some eigenvalue λ that is close to n. Then $\frac{\lambda - n}{n}$ can be used as a measure of accuracy of the estimate. If this value is large the estimate is poor and the data is just too inconsistent.

9.6. Data driven methods

9.6.1. Curve fitting

If there is a data set that can be used as a basis for the construct of the fuzzy set membership functions then a host of methods are available. The first is the common mathematical methods of curve fitting. For example, besides triangular and trapezoidal curves, Gaussian bell shaped curves with generic formula

$$Tb\,[m, s, \gamma]\,(x) = \gamma\, e^{-(x-m)^2/s^2} \tag{9.19}$$

m, s, and γ all positive, are common in mathematics. If we assume that there are n data points $\langle a_i, b_i \rangle$ then the value

$$E = \sum_{i=1}^{n} \left(b_i - \gamma\, e^{-(a_i-m)^2/s^2} \right)^2 \tag{9.20}$$

is the total squared error between the curve Tb and the given data points. This can be minimized easily and rapidly on a computer using any standard numerical technique. The solution is a the least squared error estimate of the fuzzy bell number. The result will not always be a fuzzy set for arbitrary α, β, and γ. It is then necessary to either truncate or scale the result so that it never exceeds the threshold of 1.0.

Entirely similar methods can be applied with triangular fuzzy numbers $Tr\,[a, m, b]$ and trapezoidal fuzzy numbers $Tp\,[a, l, r, b]$ to determine the values of the parameters a, b, . . . , that minimize the least squared error.

Almost any curve fitting methodology in the mathematical canon can be adapted to finding fuzzy set membership values. These include regression and Lagrangian interpolation.

9.6.2. Histograms

Sometimes it makes more sense to fit the curve to the histogram of the data set. This is the case when we do not have input-output pairs, rather we have a large sample of the input values that are not correlated with a controlled output. This is the case when no successful control, human or machine, exists.

If we have n elements $\langle a_i, b_i \rangle$ in a data set then the first step is to find $a_{\min}$ and $a_{\max}$, the minimum and maximum value obtained by the first value in the ordered pair. The next step is to divide the interval $a_{\min}$ to $a_{\max}$ into m segments, where m is usually much smaller than n. For each interval

$$[a_{j-1}, a_j] \tag{9.21}$$

where $a_0 = a_{\min}$ and for $j \in \mathbb{N}_m$

$$a_j = a_o + j \cdot \frac{a_{\max} - a_{\min}}{m} \tag{9.22}$$

we calculate the number, h_j, of points a_i that fall into the interval $[a_{j-1}, a_j]$. Let $h_{\max}$ be the maximum value obtained by data set $\{h_j\}$ and let

$$\hat{h}_j = \frac{h_j}{h_{\max}} \tag{9.23}$$

Also calculate $\hat{a}_j$, the midpoint of each interval with

$$\hat{a}_j = \frac{a_{j-1} + a_j}{2} \tag{9.24}$$

Finally we can now use the set of points $\left\langle \hat{a}_j, \hat{h}_j \right\rangle$ for curve fitting as described in the previous section.

9.6.3. Histograms for discrete data

If the data can only take on limited values, the histogram method can be simplified. Suppose that we are using observational data. For example the data might be the number of cups of coffee that employees drank on a certain day at the office. For discrete data, that can only take on a fixed set of values, the standard method of fuzzification is to use frequency data.

Suppose that the data can only take values in X with $|X| = n$. For the coffee experiment, the employees only drank 2 to 5 cups, and the number of cups is always an integer. We can then process the the data to calculate frequencies f_i for $i \in \mathbb{N}_n$. The frequency f_i is just the count of the number of times that x_i occurs in the data set d.

Example 74. Let $X = \{2, 3, 4, 5\}$. Let us convert the information in $data = [5, 3, 4, 2, 4, 3, 4, 2]$ into frequencies. The value 5 occurs once, the value 4 occurs three times, and the values 2 and 3 occur twice each. A graph of the frequency data is called a histogram. If we then divide the frequencies by the largest of the frequencies, 3, we produce a fuzzy set D. These results are summarized in Table (9.1). The graphical result is illustrated in Fig. (9.2).

x_i	f_i	$D(x_i)$
2	2	0.667
3	2	0.667
4	3	1.0
5	1	0.333

Table 9.1.: Frequency table.

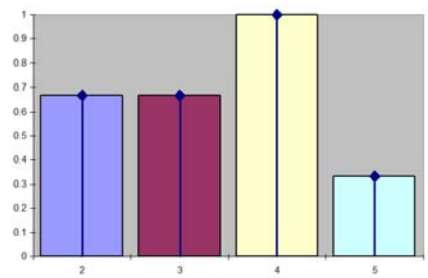


Figure 9.2.: Using a histogram to generate a fuzzy set.

9.7. Adaptive methods

The most important automatic methods for determining membership functions include:

- Artificial Neural Networks,
- Genetic Algorithms,
- Deformable Prototypes,
- Gradient Search, and
- Inductive Reasoning

One of the most common methods of calculating the fuzzy set membership functions from data is the use of neural networks. There are many advantages to this methodology. First, it makes no assumptions about the appropriate shape to fit to an input-output data set $\langle a_i, b_i \rangle$. Secondly, it functions well empirically regardless of the nature and source of the data set. Third, it can be employed on-line, or off-line, to continuously refine the membership functions in the face of new behavior of the system which one is trying to control. This adaptive behavior is particularly desirable in the construction of robust systems.

9.7.1. Neural network

A neural network is a (computer) simulation of the architecture used in the human brain. The brain of all animals in fact are made of a very large collection of processing units called neurons.. The neurons are connected together in a large network that takes in information from our senses and processes it for recognition and subsequent action.

Neural networks have the additional advantage of robustness and adaptability. A neuro–fuzzy controller can be built that changes the fuzzy set membership functions over time as conditions change. This allows a fuzzy controller for a dam to deal with drought and flood as well as the typical situation. Neural networks are so important to fuzzy applications that there is a supplamentary Chapter online at `http://duck.creighton.edu/Fuzzy/` to their description and use.

9.7.2. Genetic algorithm

Genetic algorithms are biologically inspired techniques to evolve better fuzzy sets. To accomplish this evolution the computer needs a goal. The goal in fuzzy set theory is better fuzzy logic controllers. Genetic algorithms, like neural networks, are important enough to deserve their own separate Chapter at `http://duck.creighton.edu/Fuzzy/`.

9.7.3. Deformable prototypes

An early approach by Bremermann is based on the idea of a deformable prototype. Devised initially for pattern recognition, it appears to be a potentially useful method for automatic determination of membership functions. It is based on the concept of taking an object which needs identifying and deforming it to match a prototype. The amount of matching and distortion are measured by a distortion function and a matching function to combine to give a cybernetic functional. In other words there is a combination of the matching of an object to a prototype and the distortion required to deform the prototype. Using Bremermann's notation the matching functional would be

$$\langle M\left(\Phi_i\left(p_1, p_2, \ldots, p_n\right)\right), \phi\rangle \tag{9.25}$$

where ϕ is object Φ_i is the ith prototype, and $p_1, p_2, \ldots, p_n$ are the parameters that control the distortion of the prototype. Denote

$$D\left(\Phi_i\left(p_1, p_2, \ldots, p_n\right)\right) \tag{9.26}$$

to be the distortion function then the cybernetic functional for the ith prototype can be defined by

$$\langle(f_i, \phi)\rangle = \min_{p_1, p_2, \ldots, p_n} \langle F_i\left(p_1, p_2, \ldots, p_n\right), \phi\rangle \tag{9.27}$$

where

$$\langle F_i\left(p_1, p_2, \ldots, p_n\right), \phi\rangle = \langle M\left(\Phi_i\left(p_1, p_2, \ldots, p_n\right)\right), \phi\rangle \tag{9.28}$$

$$+ cD\left(\Phi_i\left(p_1, p_2, \ldots, p_n\right)\right) \tag{9.29}$$

and c is a constant. Bremermann then goes on to explain that this cybernetic functional can be used to generate a membership function

$$\chi(\phi) = 1 - \frac{\langle f, \phi \rangle}{\max} \tag{9.30}$$

where $\max$ is the least upper bound for f. This initial suggestion seems to have been hampered by the speed of computers and techniques available at the time to carry out the optimization of the parameters. However Bremermann (1976) reports on successful implementations for pattern recognition and in particular for fuzzy sets in ECG interpretation. This does raise the question of whether, for instance, genetic algorithms may be useful in conjunction with this technique. This work still raises questions about arriving at a suitable prototype and choosing the various functions.

9.7.4. Gradient search

Based on the original work of Procyck and Mamdani (1979), Burkhardt and Bonissone (1992) use gradient search methods to tune fuzzy knowledge bases. They fine tune the rules (see Chapter (17)) and also the membership function for a fuzzy logic controller. They assume first, as with most fuzzy logic controllers, that the membership functions are triangular and then they use gradient search to determine the optimal scaling factor for the base of the triangles. This optimization is carried out on an application of fuzzy logic control to the cart-pole system (this is the standard example, and will be explained in detail in Chapter (18)) where the goal is to maintain the time the pole is vertical on the cart while reducing overshoot (the pole falls off) and steady state error (the pole wobbles a lot). They present a variety of results and draw the conclusion that this approach out-performed a simple controller.

9.7.5. Inductive reasoning

The approach adopted by Kim and Russell (1993) is to use inductive reasoning to generate the membership functions and the rules. They assume they have no information other than a set of data. The approach is to partition a set of data into classes based on minimizing the entropy. The entropy where only one outcome is true is the expected value of the information contained in the data set and is given by

$$S = k\Sigma_{i=1}^{N}[p_i \ln p_i + (1 - p_i)\ln(1 - p_i)] \tag{9.31}$$

where k is an arbitrary constant, the probability of the i-th sample to be true is p_i, and N is the number of samples. In their work they give a two class example where by iteratively partitioning the space they calculate an estimate for the entropy which leaves them with points in the region that are then used to determine triangular membership functions. This approach suffers from the fact that there is no way of knowing whether the membership functions are realistic and that the sets obtained are triangular.

9.8. Notes

Early survey work on fuzzification is presented in Turksen (1991). Klir and Yuan (1996) devotes a Chapter of their book to the construction of fuzzy membership function. More references to fuzzy set construction are in the Notes section of the Neural Net chapter at `http://duck.creighton.edu/Fuzzy/`.

9.9. Homework

1. Obtain some color samples sheets from a paint store or department store that sells paint. You could also use a color printer to produce color swatches. Number all the color sheets on the back and present them in random order to an audience. Give them a sheet of paper with the numbers in a list. For each sheet have them answer the question "How red is this sheet?" Having the audience rate the redness using values in the 0-10 range since this data is easy to process. You can choose some other color than red of course. If you use a color printer then each color sheet has a redness value that can be read from the computer. For paint sheets a scanner can provide an absolute redness. Consider this data as a basis for construction of a fuzzy set for redness perception. How well does the audience do when compared with the computer or scanners redness rating.

2. Apply the indirect method with one expert for the data in the relative wealth table to determine a fuzzy set that describes the degree of membership of each person in the fuzzy set rich.

Wealth	Alex	Bela	Cora	Dean	Eden	Fara
Alex	1	3	4	2	7	5
Bela	$\frac{1}{3}$	1	1	3	5	6
Cora	$\frac{1}{4}$	1	1	5	4	3
Dean	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{5}$	1	$\frac{1}{2}$	$\frac{1}{2}$
Eden	$\frac{1}{7}$	$\frac{1}{5}$	$\frac{1}{4}$	2	1	$\frac{1}{3}$
Fara	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{3}$	2	3	1

3. Use the following data set to construct fuzzy sets. Assume that the x values are in the range $0-10$ and that the y values are in the unit interval.

   $\langle 9, .21\rangle, \langle 2, .5\rangle, \langle 10, 0.01\rangle, \langle 6, .66\rangle, \langle 8, .44\rangle, \langle 0, 0.1\rangle, \langle 7, .55\rangle, \langle 1, .15\rangle, \langle 3, .6\rangle, \langle 4, .8\rangle, \langle 5, 1\rangle.$

4. Subjectively construct fuzzy sets that represent how you personally perceive age. Construct fuzzy sets for the four linguistic terms of Ch. (16), young, old, infant, adolescent that represents your own opinion.

   For the following questions, assume that we have given the question sheet in Fig. **??** to a group of acquaintances and tabulated the data.

5. Use the answer to questions 1-8 of the *Questionnaire* **??** as a data set to construct a fuzzy set young. What operator very seems to produce very young from young?

6. Use the answer to the questions 9-20 that represent vertical data to construct fuzzy sets for the four linguistic terms of Ch. (16) young, old, infant, adolescent.

7. How do the results of Question 4. compare with the results in the previous question.

8. Use the answer to the questions 9-20 that represent horizontal data to construct fuzzy sets for the four linguistic terms of Ch. (16) young, old, infant, adolescent.

9. How do the results of Question 4 compare with the results in the previous question.

10. Construct a Questionnaire to gather data about the perception of height. What linguistic terms would you use to categorize human height. Administer the Questionnaire and build fuzzy sets from the resulting data.

10. Fuzzy Clustering

10.1. Introduction

The modern world is awash in data. This is one of the consequences of the computer revolution. The computer allows for the collection and storage of vast amounts of data. It is becoming apparent that we have more data than understanding.

This chapter focuses on clustering and will present two algorithms used to cluster data. Another clustering method is presented online in the Neural Network Chapter `http://duck.creighton.edu/Fuzzy/`.

> Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic concepts and methodologies slow to occur. — Jain et al. (1999).

10.2. Clustering

Suppose a company obtains an aerial photograph of a geographic area being farmed by one of its subdivisions. It wants to analyze this photograph to determine what part of the plot represents corn, what part represents soybeans, and what part of the image has no crops at all. It would also like some data on what percent of the crop is ripe so that an optimum harvest time can be determined.

This is the type of problem that clustering algorithms are used to solve. By clustering the data into a collection of clusters we hope that all the corn data will go into one cluster, all the soybean data will go into a second cluster, and that the other objects present, rocks, trees, grasses, etc. will be grouped into other clusters.

When we examine the cluster for corn, for example, we can also hope that the shape of the cluster will allow the observer to infer something about the ripeness. A tight cluster means all the corn is in the same state and if some are ripe all are ripe. If the cluster is diffuse then it would indicate that different parts of the cultivation area are in different states of ripeness.

Many other domains use clustering algorithms to extract information by grouping a large collection of objects; such as the medical sciences (clustering of symptoms yields information about the diseases), the earth sciences (to determine land use patterns), and marketing research (to target advertisements and promotions towards specific consumer groups that have a high probability of responding to them).

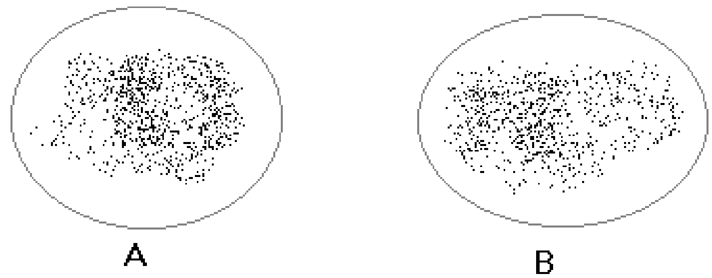


Figure 10.1.: Ideal data seperated into two discrete optimal clusters A and B.

If the data is ideal, such as that pictures in figure (10.1) then there exists traditional clustering methods, such as k–meansBandemer and Nather (1992) and ISODATABall and Hall (1965), (Iterative Self-Organizing Data Analysis Technique), that will partition the data into two sets A and B that have no overlap. Both k–means and ISODATA are popular self–organizing algorithms that seek to minimize the distance of the data items to cluster prototypes (usually the center).

However, most data is not like the kind in figure (10.1). There are all kinds of reasons for the data to be difficult to analyze. These problems include:

outliers data points that are inconsistently large or small,

bad data transcription or measurement error, and

inconsistent data information gathered over time may be dependant on an unmeasured parameter such as temperature.

The following sections will present the k–means clustering technique and the fuzzy c–means algorithm.

10.3. k–means

Figure (10.2) is a scatter plot of a set of n ordered pairs or points $x_j = \langle x_{j1}, x_{j2}\rangle$ with $j \in \mathbb{N}_n$. Clustering asks the question "Can we partition the set of n data points x_i into c subsets where each subset contains 'similar' points?"

A partition of a set is a collection of nonempty, disjoint subsets whose union is the original set. If $X = \{a, b, c\}$ then a partition into two subsets, $c = 2$, would be $S = \{\{b\}, \{a, c\}\}$ because $\{b\} \cap \{a, c\} = \emptyset$ and $\{b\} \cup \{a, c\} = X$. The pieces $\{b\}$ and $\{a, c\}$ are called clusters.

Figure (10.3) is a scatter plot of three points

$$
\begin{aligned}
a &= \langle 1, 2\rangle\,, \\
b &= \langle 3, 5\rangle\,, \text{ and} \\
c &= \langle 2, 4\rangle\,.
\end{aligned}
\tag{10.1}
$$

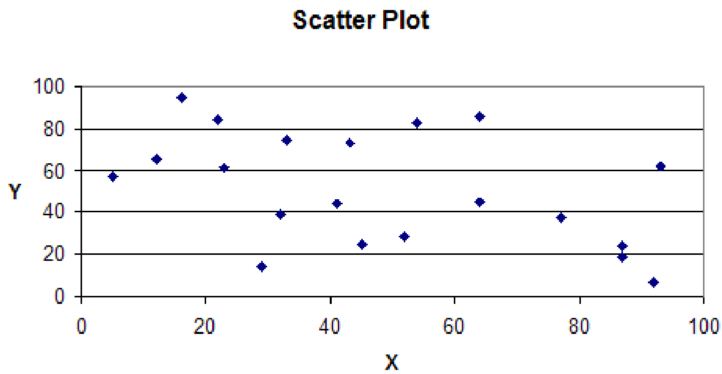


Figure 10.2.: A scatter plot of 20 pairs of $< x, y >$ values.

We want to partition the set $X = \{a, b, c\}$ into two subsets of similar points. What does it mean for two points to be similar? In this case let us define two points to be similar if they are close in geometric distance. The closer the two points are to each other the more similar they are to each other.

We can make a simple table that shows the classification of the three points. Let C_1 and C_2 be cluster one and cluster two respectively. Then a three by two table can represent the present classification of the three objects. For example Table (10.1) presents the partition $S = \{\{c\}, \{a, b\}\}$. A 1 indicates that the point that heads that column is in the cluster that heads that row. An 0 indicates the opposite, that the point that heads that column is not in the cluster that heads that row.

Is the partition in table 10.1 a good partition? Is there another partition where the degree of dissimilarity is less? It is possible to search exhaustively through every possible partition of $X = \{a, b, c\}$ but if we were dealing with Figure (10.2) then the number of different partitions is one of those combinatorial problems that grows too large too fast to be dealt with feasibly. We need a simpler and more efficient method to deal with large data sets.

The algorithm k–means is the name of the method developed to iteratively cluster a set of data points. Its methodology is fairly simple. The first step is to decide how many clusters one wants the algorithm to generate. Right now we will assume that we wish to cluster the data into two classes, C_1 and C_2. The second step is to randomly assign the data points to clusters, making sure that no clusters is initially empty. This means you have to go through a Table like Table (10.1) and put a 1 somewhere in each column, filling in the rest of the column with 0s, and make sure that each row also has at least one 1.

${}^0S$	a	b	c
C_1	0	0	1
C_2	1	1	0

Table 10.1.: A partition of the data set $X = \{x, y, z\}$ into two clusters C_1 and C_2.

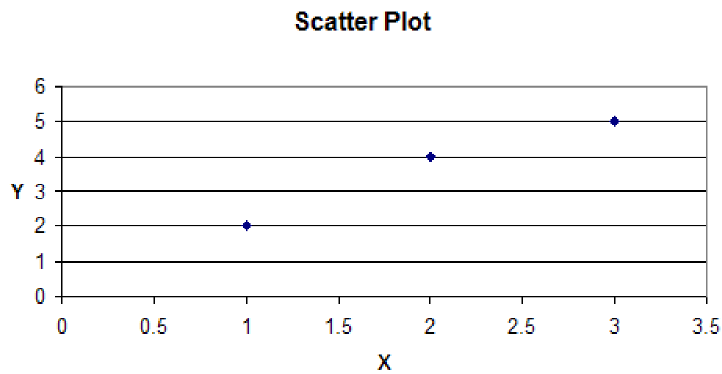


Figure 10.3.: A scatter plot of the three points a, b, and c.

We are now finished with the initialization steps.

The third step is to find the central point geometrically of each cluster. It turns out that this value is easy to find because it is simply the centroid, or center of gravity (this value will also be important in Chapter 17 on fuzzy control). Let c_1 be the center of cluster C_1 and let c_2 be the center of cluster C_2.

Step four is to go through the data points and calculate the distance of each point to the centroids c_1 and c_2. If a point is closer to the centroid of a cluster it is not in than it is to the centroid of the cluster it is in it has been incorrectly classified and should be shifted from its original cluster to the new cluster whose centroid it is nearer. This is step five where we reclassify the data points to produce a new partition.

Table (10.2) shows the centroids of the various clusters and the distances of the various points to the centroids of the two clusters. In this table i is the cluster number, c_i is the clusters center and $\delta(a, c_i)$ is the distance from the center to the data point a and the numerical values for a, b, and c are given in Eq. (10.1).

i	Cluster	centroid c_i	$\delta(a, c_i)$	$\delta(b, c_i)$	$\delta(c, c_i)$
1	$C_1 = \{c\}$	$c_1 = \langle 2, 4\rangle$	2.236	1.414	0.000
2	$C_2 = \{a, b\}$	$c_2 = \langle 2, 3.5\rangle$	1.803	1.803	0.500

Table 10.2.: Distances of the data points $\{x, y, z\}$ to the centroids of the cluster centers c_1 and c_2.

Let us examine the meaning of Tables (10.1) and (10.2). From Table (10.1) we see that point a is in cluster C_2 and from Table (10.2) it is apparent that a is closer to the center of C_2 than it is to the center of cluster C_1, i.e.,

$$\delta(a, c_2) = 1.803 < \delta(a, c_1) = 2.236.$$

The same is not true for point b. If we examine the numerical distances $\delta(b, c_1) = 1.414$ and $\delta(b, c_2) = 1.803$ we see that point b is mis-classified. It is presently classified in C_2 but in fact it is in fact closer to the center of cluster C_1. It is mis–classified. Point

b should be reclassified into cluster C_1. Finally point c is correctly classified, it is closest (distance zero) from the center of its cluster C_1. If we reclassify b we get the Table (10.3).

${}^1S$	a	b	c
C_1	0	1	1
C_2	1	0	0

Table 10.3.: A partition of the data set $X = \{x, y, z\}$ into two clusters C_1 and C_2.

We now loop. Reiterate the steps three through five, which consist of; (3) calculating the centroid, (4) determining the distances from point to centroid and then (5) reclassifying the data points. Loop until

1. a) no points have changed classification, or
   b) some predefined limit on the number of iterations has been exceeded.

10.3.1. The k–means algorithm

To try to make the algorithm a little clearer we state that:

- There are c clusters C and the index i is used to indicate a specific cluster. Each cluster has a prototype value c_i, typically the centroid..
- There are n data points x and the index j is used to indicate a specific point.
- The data is q dimensional and the index k is used to indicate a specific dimension.
- The current iteration number is t and the current partition matrix (solution) is ${}^tS$.
- The algorithm loops until the current partition matrix ${}^tS$ is identical to the previous partition matrix ${}^{t-1}S$ or until $t_{\max}$ iterations have been performed.

The complete algorithm, to classify n points into c classes is:

1. Randomly assign the n points to the c classes making sure that each cluster contains at least one data point. It is easiest to think of the result as a matrix ${}^0S_{cn}$ with c rows and n columns with ${}^0s_{i\,j}$ being the value in the ith row and the jth column of the matrix ${}^0S_{cn}$. The value ${}^0s_{i\,j}$ must be zero or one so we know that ${}^0s_{i\,j} \geq 0$. If ${}^0s_{ij}$ is one then we interpret this as saying data point j is in cluster i. We require that each column in the initial matrix have only a single 1 in it, a requirement which can be expressed by the constraint

   $$\sum_{i=1}^{c} {}^0s_{i\,j} = 1 \text{ for all } 1 \leq j \leq n.$$

   Additionally, it is required that each cluster be non–empty so that values of $s_{i\,j}$ for each row must sum to be greater than or equal to one, i.e.,

   $$\sum_{j=1}^{n} {}^0s_{i\,j} \geq 1 \text{ for all } 1 \leq i \leq c.$$

Finally set t the iteration counter to one, $t = 1$.

2. Calculate ${}^t c_i$ the centroid of cluster ${}^t C_i$ by calculating the center of gravity in each of its dimensions. We have used two dimensions in our example but there is no difficulty in generalizing to q dimensions. For $1 \le k \le q$

$$ {}^t c_{i\,k} = \frac{\sum_{j=1}^{n} {}^{t-1}s_{i\,j}\, x_{jk}}{\sum_{j=1}^{n} {}^{t-1}s_{i\,j}}. $$

   and the entire centroid ${}^t c_i$ is the ordered q–tuple $\langle {}^t c_{i1}, {}^t c_{i2}, ..., {}^t c_{iq} \rangle$

3. For each centroid ${}^t c_i$ calculate the distance of each data point to the centroid of the various cluster. Typically the distance is the euclidean distance from the data point x_j to centroid c_i

$$ \delta_{i\,j} = \delta({}^t c_i, x_j) = \sqrt{\sum_{k=1}^{q} \left({}^t c_{i\,k} - x_{j\,k}\right)^2} $$

4. Create a new tableau ${}^t S_{n\,c}$ with $s_{i\,j} = 1$ if and only if the jth cluster center is the closest center to the data point i. That is, if $\delta_{i\,j} \le \delta_{i\,j'}$ for all $1 \le j' \le c$ then $s_{ij} = 1$ otherwise $s_{ij} = 0$. In other words classify x in the cluster with the nearest centroid. If two centroids are equally close put it in the emptier cluster. If all nearest clusters are equally empty choose one at random. Make sure that all cluster are non-empty. If any cluster is empty put in the nearest data value in terms of distance. In reality this is not a problem since the central points should be closer to their own centroid than they are to any other clusters centroid. Finally set $t = t + 1$. (We also note that in this algorithm we can actually use the square of the distance $\delta_{i\,j}^2 = \sum_{k=1}^{q} \left({}^t c_{i\,k} - x_{j\,k}\right)^2$ in both Steps **3** and **4** which will speed up the execution of the algorithm.)

5. Repeat steps 2-4 until a no data changes clusters or until a fixed number of iterations is exceeded.

If we were to write a computer program to accomplish this task we would use the algorithm that follows. For this purpose we note that in the program we need only keep track of the current classification matrix $currentPartition$ and the previous classification matrix $previousPartition$. This produces Algorithm (10.1).

The ISODATA algorithm is essentially the k–means algorithm with some further refinements for the splitting and merging of clusters. Three parameters are required for ISODATA, two splitting parameter s_s for size and s_d for distance and one merging parameter σ_0 for spread. Clusters are merged when they are small or they are close. This occurs when either the number of members in a cluster is less than the threshold value s_s, or if the centers of two clusters are closer than a certain threshold s_d. Clusters are split into two different clusters if they are large and spread out. This occurs when the cluster standard deviation exceeds a predefined value σ_0 and the number of members is twice the threshold for the minimum number of members s_s.

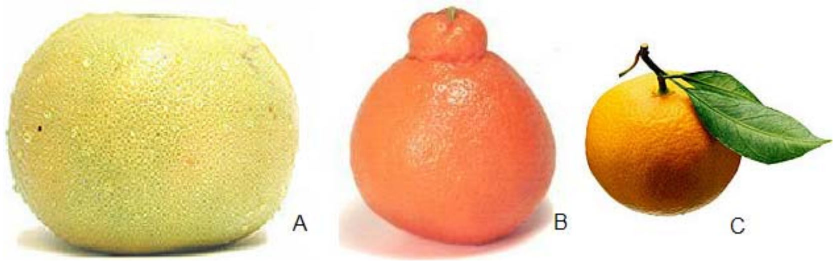


Figure 10.4.: Three citrus fruits.

10.4. Problems with k–means

The major problem associated with k–means is one that is common to many of the problems examined in this book. Suppose we have three points $a = \langle 1, 1\rangle$, $b = \langle 2, 2\rangle$ and $c = \langle 3, 3\rangle$ and we want to produce two cluster. A casual inspection will show that point b is equidistant from points a and c. Which class should b belong to? k–means is one of those Aristotelian, Boolean, **yes-no**, **0–1** mechanisms. It must classify b in either C_1 or C_2 and b obviously has the same distance to a as it has to c. To partition the data into the sets $\{a, c\}$ and $\{b\}$ is nonsensical; certainly b is closer to either a or c than a is to c. The Algorithm (10.1) will eventually terminate, once it puts b into a cluster with a or c the program will stop. If we run the program again we can get a different answer, in fact, repeated runs should cluster a and b fifty percent of the time and cluster b and c the other fifty percent. This result would seem to indicate the methodology is not sure where to place b, and if the algorithm is not sure, why should we be! If, however, we allow the partial placement of b with both a and c we have a fuzzy situation.

This is not an artificial situation. Consider the three citrus fruits in figure 10.4. If we have to put A, B, and C into two bins and not leave a bin empty we must group two of these fruits together. Now A and C are both round but they are very different in size and color. If we put A and B together or B and C together that would seem more reasonable. However A is a grapefruit, B is a tangelo, and C is a tangerine. The tangelo is a crossbreed of a tangerine and a grapefruit. It is therefore exactly 50% tangerine and 50%grapefruit.

A second problem with k–means (and for that matter many clustering algorithms) is in dealing with outliers. An outlier is a data item that does not seem to fit the data set. A sixty-six foot tall human being would be an outlier. We would be suspicious that the data entry procedure involved a stuck key and that this is supposed to be a six foot tall human being. However, if we have no way of verifying that the outlier is indeed inaccurate then the outlier tends to throw off the accuracy of the clustering.

When both of these problems are taken into account k–means can be adapted into Keller's [1981] fuzzy c–means algorithm.

10.5. Fuzzy c-means

With a background in fuzzy set theory we can see immediately that if we replace Aristotelian reasoning with Zadehian reasoning, that there is no reason not to classify b as somewhat similar to both a and c. To put it another way k–means requires the classification index $s_{i\,j}$ of the membership of data item x_j in cluster C_i to be zero or one. Let us relax this requirement so that it is allowed that $0 \leq s_{i\,j} \leq 1$ and keep everything else in the k–means algorithm as identical to the original method as is feasible. There are only two other significant change in the algorithm. The nonempty requirement for a fuzzy set is not as stringent as for a crisp set. A crisp set is not empty if it contains at least one element. A fuzzy set is not empty if it contains any element to any non-zero degree. The second difference between fuzzy c–means and k–means is the introduction of a weighing parameter m the allows fuzzy c-means to deal with outliers successfully. Fuzzy k–means, also called fuzzy c-means,Bezdek (1981) is a fast and robust clustering algorithm.

10.5.1. The Fuzzy k–means algorithm

Let us restate the conventions of k–means.

- There are c clusters C and the index i is used to indicate a specific cluster. Each cluster has a prototype value c_i.
- There are n data points x and the index j is used to indicate a specific point.
- The data is q dimensional and the index k is used to indicate a specific dimension.
- The current iteration number is t and the current partition matrix (solution) is ${}^tS$.

The complete fuzzy **k–means** algorithm, to classify n points into c classes is:

1. Randomly assign the n points to the c classes making sure that each cluster contains at least one data point. The result is a matrix ${}^0S_{cn}$ with c rows and n columns with ${}^0s_{i\,j}$ being the value in the ith row and the jth column of the matrix ${}^0S_{cn}$. The initial value of ${}^0s_{i\,j}$ is still zero or one, but throughout the duration of the algorithm we constrain ${}^0s_{i\,j}$ to be in the unit interval ${}^0s_{i\,j} \in [0, 1]$. The interpretation of ${}^0s_{i\,j}$ is the degree of membership of data point j in cluster i. We will still require that each column sum to one, but as the algorithm iterates the values of ${}^0s_{i\,j}$ will move through the unit interval, i.e., become fuzzy membership grades. We still have the constraint

$$\sum_{i=1}^{c} {}^0s_{i\,j} = 1 \text{ for all } 1 \leq j \leq n.$$

   And we also require that each cluster be non–empty so that values of $s_{i\,j}$ for each row must be greater than zero, i.e.,

$$\sum_{j=1}^{n} {}^0s_{i\,j} > 0 \text{ for all } 1 \leq i \leq c.$$

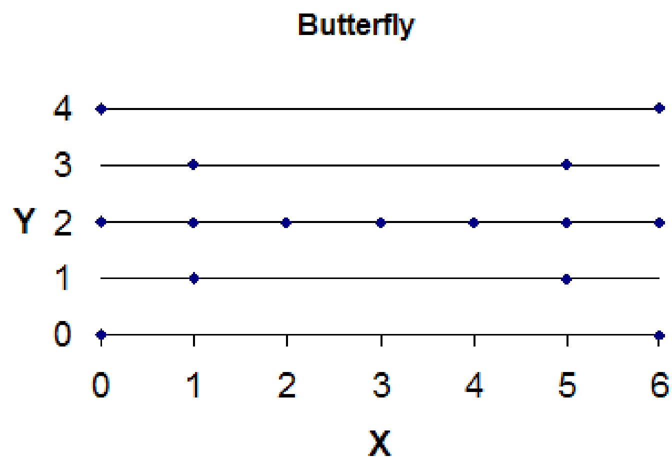


Figure 10.5.: An example of a data set that a regular k-means program would find hard to cluster. The point $\langle 3, 2\rangle$ has some degree of membership in both the left and right cluster.

Finally set the iteration counter t to one, $t = 1$.

2. Calculate ${}^tc_i$ the weighted centroid of cluster ${}^tC_i$ by calculating the center of gravity in each of its dimensions. However a weighing parameter m is introduced that helps controls noise (outliers and bad values) in the data. We have used two dimensions in our example but there is no difficulty in generalizing to q dimensions. For $1 \leq k \leq q$

$$ {}^tc_{ik} = \frac{\sum_{j=1}^{n} {}^{t-1}s_{i\,j}^{m}\; x_{jk}}{\sum_{j=1}^{n} {}^{t-1}s_{i\,j}^{m}} $$

and the entire centroid c_i is the ordered q–tuple $\langle {}^tc_{i1}, {}^tc_{i2}, ..., {}^tc_{iq}\rangle$.

3. For each centroid ${}^tc_i$ calculate the similarity to the data points x_j. The formula with similarity measured by euclidean distance from the data point x_j to centroid ${}^tc_i$ is:

$$ \delta_{i\,j} = \delta(c_i, x_j) = \sqrt{\sum_{k=1}^{q} \left({}^tc_{ik} - x_{jk}\right)^2} $$

4. Create a new tableau ${}^tS_{cn}$ with

$$ {}^ts_{i\,j} = \frac{1}{\sum_{l=1}^{c} \left(\frac{\delta_{i\,j}^2}{\delta_{lj}^2}\right)^{\frac{1}{m-1}}} $$

unless $\delta_{i\,j} = 0$ for some $i\,j$. If $\delta_{i\,j'} = 0$ for some j' then set ${}^{t}s_{i\,j'} = 1$ and set ${}^{t}s_{i\,j} = 0$ for all $j \neq j'$. In other words classify x_j in the cluster ${}^{t}C_i$ with membership degree ${}^{t}s_{i\,j}$ so that x_j belongs to various clusters to varying degrees, but also so that x_j has stronger membership in clusters that it is closer to. Then set $t = t + 1$.

5. Calculate the distance $\Delta^t S$, between ${}^{t-1}S$ and ${}^{t}S$. Either the Hamming distance or the Euclidean distance is acceptable here.

$$\Delta^t S = \Delta\left({}^{t-1}S, {}^{t}S\right) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left\|{}^{t-1}s_{i\,j} - {}^{t}s_{i\,j}\right\|.$$

If $\Delta^t S$ is less than some stopping distance HALT, else repeat Steps 3-5 until $\Delta^t S$ becomes acceptable or until a t exceeds some fixed number of iterations.

The parameter m is an exponential weighing that reduces the effect of noise. It ensures that close points have a greater effect upon the center of gravity and subsequent membership grades in the clusters than distant points with small but significant membership grades.

The butterfly data that produces Fig. (10.5) looks like:

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
x_1	0	0	0	1	1	1	2	3	4	5	5	5	6	6	6
x_2	0	2	4	1	2	3	2	2	2	1	2	3	0	2	4

Table 10.4.: The Butterfly data set..

When we run the crisp k–means algorithm on this data we get the following membership distribution. Note that b_8 which is in the middle of the data ends up in cluster C_2.

	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
C_1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
C_2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0

Table 10.5.: A partition of the Butterfly data set two fuzzy clusters C_1 and C_2 when m=2.0.

When we run the fuzzy k–means algorithm on this data with $m = 1.5$ we get the following table of membership grades as the final result.

10.6. Comments on clustering algorithms

The algorithms for k–means and fuzzy c–means clustering are presented in this chapter. Jain [1999] is a good overview of the various clustering algorithms, their strengths, and their weaknesses. For example, rate of convergence is very important for large data sets. For very large data sets an algorithm that gives good clusters fast may be preferable to an algorithm that gives more precise results but at a much later

	b_1	b_2	b_3	b_4	b_5
C_1	0.145	0.036	0.145	0.055	0.000
C_2	0.855	0.964	0.855	0.945	1.000
	b_6	b_7	b_8	b_9	b_{10}
C_1	0.055	0.103	0.500	0.897	0.945
C_2	0.945	0.897	0.500	0.103	0.055
	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
C_1	1.000	0.855	0.855	0.964	0.855
C_2	0.000	0.055	0.145	0.036	0.145

Table 10.6.: A partition of the Butterfly data set two fuzzy clusters C_1 and C_2 when m=21.5.

date. This would be true for a investment company that needed to make real time decisions.

In both k–means and fuzzy c–means algorithms, a fixed number of clusters is specified before the data is processed. Variations of these, and other clustering algorithms, might allow for the splitting of large diffuse clusters and/or the merging of close sparsely populated clusters. An alternative approach to the problem of cluster numbers is to put the algorithm inside an outer loop that increments the number of clusters, $c = 2, 3, \dots$, to find the best cluster size. For instance, we might judge the overall goodness of a result for c clusters based on the average cluster variance. The variance would be the average distance of every point in the cluster from the centroid.

Another aspect of k–means and fuzzy c–means is that they cluster the data about points. A natural variation would be to cluster the data about lines, circles or any quadratic curve. An application area where this approach might be useful is image processing where the discovery of lines in the image is called edge finding and is a major task in image recognition.

10.7. Notes

The paper Jain et al. (1999) is an excellent overview of clustering algorithms. The books Bezdek (1981), Kandel (1982), Bezdek and Pal (1992), Bezdek et al. (1999), and Bandemer and Nather (1992) provide wonderful resources, especially bibliographies, for further research in the area of pattern recognition.

10.8. Homework

Let us define the data sets

$$A = \{\langle 0,4\rangle, \langle 1,3\rangle, \langle 2,5\rangle\}, \tag{10.2}$$

$$B = \{\langle 0,4\rangle, \langle 0,8\rangle, \langle 1,5\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \langle 2,5\rangle, \langle 2,6\rangle\}, \tag{10.3}$$

and

$$C = \{\langle 1,0,4\rangle, \langle 0,2,8\rangle, \langle 1,9,5\rangle, \langle 3,1,2\rangle, \langle 6,2,1\rangle, \langle 2,5,7\rangle, \langle 8,2,6\rangle, \langle 3,6,5\rangle\}. \tag{10.4}$$

1. Use k–means to cluster data set A by hand into two partitions.
2. Use k–means to cluster data set B by hand into two partitions.
3. Use k–means to cluster data set C into three partitions using a spreadsheet such as EXCEL.
4. Use fuzzy c–means to cluster data set A by hand into two partitions with $m = 2.0$ and with $m = 1.5$. How do these results compare with the results of problem 1.
5. Use fuzzy c–means to cluster data set B into two partitions. Experiment with different values of m. Graph the results. How do these results compare with the results of problem 2.
6. Use fuzzy c–means to cluster data set C into three partitions using a spreadsheet such as EXCEL. How do these results compare with the results of problem 3.
7. Write a program to process the butterfly data set.
8. How many clusters should we use for the data in Figure (10.2).
9. What is the result of clustering the data in Figure (10.2).
10. Gather data from your class about three preferences such as Favorite Day of the Week, Favorite Cuisine, and Preferred Classtime:

Day		**Cuisine**		**Time**	
Sunday	1	American	1	Morning	1
Monday	2	Chinese	2	Afternoon	2
Tuesday	3	Italian	3	Evening	3
Wednesday	4	Mexican	4	Weekend	4
Thursday	5	Other	5		
Friday	6				
Saturday	7				

    Thus a data item of $W = \langle 5,2,2\rangle$ indicates that person W likes Thursday, Chinese food, and Afternoon classes. Cluster the students in the class into two or three clusters using k–means and fuzzy c–means. Explain and compare the results.

11. Explain several applications of clustering in the biological sciences.
12. Market data is the kind of information collected at grocery stores concerning what customers buy. How can clustering techniques aid in understanding and interpreting market data.

Algorithm 10.1 k–means

Require: Initialization
  c – the number of clusters
  n – the number of data points
  q – the dimension of the data points
  $t_{\max}$ – the maximum number of iterations
  Input the data matrix x
  repeat
    for $i = 1$ to c **do**
      for $j = 1$ to n **do**
        $currentPartition[i, j] \Leftarrow$ rand$\%2$
        {rand returns integer and % is modulus operator}
      end for
    end for
  until $\sum_{i=1}^{c} currentPartition[i, j] = 1$ for all $1 \leq j \leq n$ and $\sum_{j=1}^{n} currentPartition[i, j] \geq 1$ for all $1 \leq i \leq c$
  repeat
    $previousPartition \Leftarrow currentPartition$
    for $i = 1$ to c **do**
      for $k = 1$ to q **do**
        $c[i, k] \Leftarrow \frac{\sum_{j=1}^{n} currentPartition[i,j]\, x[j,k]}{\sum_{j=1}^{n} currentPartition[i,j]}$
      end for
    end for
    for $i = 1$ to c **do**
      for $j = 1$ to n **do**
        $\delta[i, j] \Leftarrow \sqrt{\sum_{k=1}^{q} \left(c[i, k] - x[j, k]\right)^2}$
        {Calculate distance from data point to centroid}
      end for
    end for
    for $i = 1$ to c **do**
      $\delta \min[i] \Leftarrow \min_{i=1}^{c} \delta[i, j]$
      for $j = 1$ to n **do**
        if $d[i, j] = \delta \min[i]$ **then**
          $currentPartition[i, j] \Leftarrow 1$
        else
          $currentPartition[i, j] \Leftarrow 0$
        end if
        {Put each point in the closest partition}
        {If two centroids are equally close put it in the emptier cluster}
      end for
    end for
    {If a cluster is empty move in the nearest point}
    $t \Leftarrow t + 1$.
  until $t = t_{max}$ or $currentPartition = previosPartiton$

Algorithm 10.2 fuzzy c–means

Require: Initialization

c – the number of clusters
n – the number of data points
q – the dimension of the data points
m – the outlier weight parameter
$e_{\max}$ – the stopping distance
$t_{\max}$ – the maximum number of iterations

```
Input the data matrix x
repeat
  for i = 1 to c do
    for j = 1 to n do
      currentPartition[i, j] ⇐ rand mod 2
      {rand returns a random integer and mod is the modulus operator}
    end for
  end for
until Σ_{i=1}^{c} currentPartition[i, j] = 1 for all 1 ≤ j ≤ n and
Σ_{j=1}^{n} currentPartition[i, j] ≥ 1 for all 1 ≤ i ≤ c
repeat
  previousPartition ⇐ currentPartition
  for i = 1 to c do
    for k = 1 to q do
      c[i, k] ⇐ Σ_{j=1}^{n} (currentPartition[i,j] ^m)*x[j,k] / Σ_{j=1}^{n} (currentPartition[i,j] ^m)
    end for
  end for
  for i = 1 to c do
    for j = 1 to n do
      δ[i, j] ⇐ δ(c[i], x[j]) ⇐ sqrt( Σ_{k=1}^{q} (c[i, k] − x[j, k])^2 )
      {Calculate distance from data point to centroid}
    end for
  end for
  for i = 1 to c do
    if d[i, j] ≠ 0 then
      for j = 1 to n do
        currentPartition[i, j] ⇐ 1 / Σ_{l=1}^{c} (δ^2_{ij} / δ^2_{lj})^{1/(m−1)}
      end for
    else
      for j = 1 to n do
        currentPartition[i, j] ⇐ δ[i, j] = 0
        {δ[i, j] = 0 is boolean expression}
      end for
    end if
  end for
  {If a cluster is empty move in the nearest point}
  t ⇐ t + 1
  e ⇐ Σ_{i=1}^{c} Σ_{j=1}^{n} sqrt( (currentPartition[i, j] − previousPartition[i, j])^2 )
until t = t_max or e < e_max
```

Part III.

Evidence Theory

11. Rough Set Theory

11.1. Introduction

A recent addition to the mathematical frameworks that deals with uncertainty is rough set theory Pawlak (1991, 2001). In Chapter 8 we noted that an equivalence relation on set induces a partition of the set, and that a partition of a set can be used to define an equivalence relation. The basis of rough set theory (**RST**) is a an equivalence relation called indiscernability. Based upon this partition, rough approximations, both upper and lower, of an arbitrary subset of the universe $\mathcal{U}$ can be formed. Rough set theory is heavily used in Knowledge Discovery in Databases, or **KDD**. The basic concepts of **RST** are:

- Indiscernibility
- Set Approximation
- Reducts and Core
- Rough
- Membership Dependency of Attributes
- Information/Decision Systems (Tables)

According to Pawlak, rough sets are used to model vagueness. Vagueness is not allowed in mathematics where everything is defined precisely and calculated exactly, but vagueness is something computer scientists and philosophers deal with routinely. Pawlak also remarks that rough set theory clearly distinguishes between two very important concepts, vagueness and uncertainty, very often confused in the **AI** literature. Vagueness is the property of sets and can be described by approximations, whereas uncertainty is the property of elements of a set and can expressed by the rough membership function.

11.2. Basic definitions

Let R be any indiscernability (equivalence) relation on U and let $x \subseteq U$. Let $[x]_R$ be the equivalence class of x under the relation R, thus $[x]_R = \{y \mid x\,R\,y\}$. In rough set theory these equivalence classes are also called *granules*. Often, in **RST** the granule (equivalence class) of $x \in \mathcal{U}$ under the relation R is denoted $R(x)$. We now define the upper and lower approximations of a subset $X \subseteq \mathcal{U}$ determined by the granules of R.

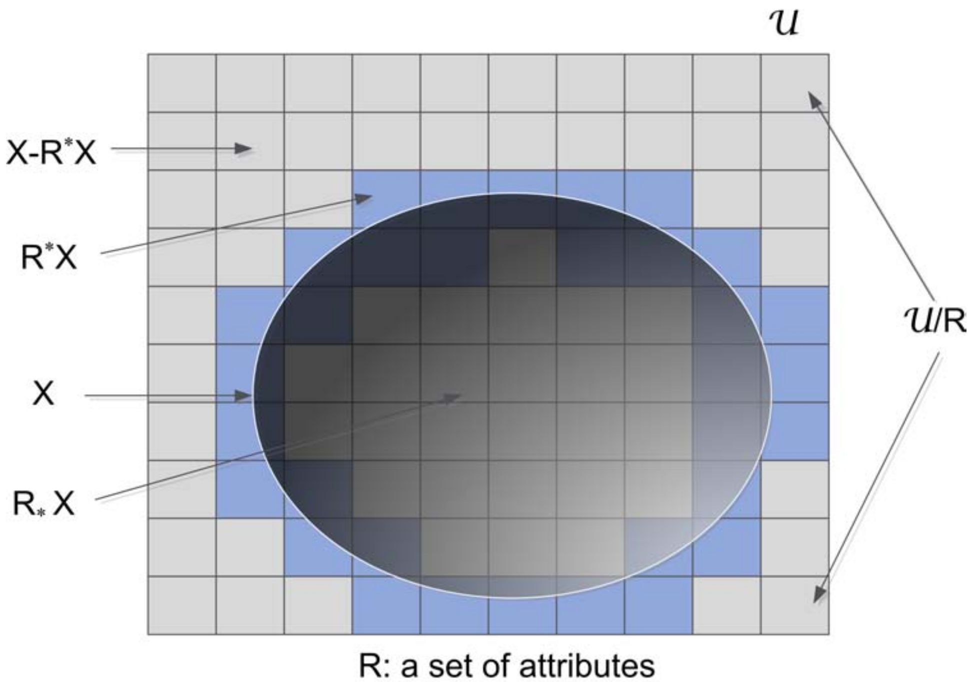


Figure 11.1.: A rough set.

Definition 58 (R–lower approximation)**.** R_*X is the R-lower approximation of X and is defined as

$$R_*X = \bigcup_{[x]_R \subseteq X} [x]_R$$

equivalently, we could use the definition

$$R_*X = \{x \mid [x]_R \subseteq X\} \ .$$

Definition 59 (R–upper approximation)**.** R^*X is the R-upper approximation of X and is defined as

$$R^*X = \bigcup_{[x]_R \cap X \neq \emptyset} [x]_R$$

equivalently, we could use the definition

$$R^*X = \{x \mid [x]_R \cap X \neq \emptyset\} \ .$$

The boundary region of X are those elements in the upper approximation that are not in the lower approximation.

Definition 60 (R–boundary region)**.** Those objects that we cannot decisively classify into X via R are called the R–boundary region of X. It is denoted $BN_R(X)$ and,

$$BN_R X = R^*X - R_*X$$

Definition 61 (R–outside region)**.** Those objects that are certainty classified as not belonging to X make up the R–outside region of X. Mathematically the R–outside region of X is equal to $U - R^*X$.

Example 75. Let $U = \{a, b, c, d, e, f\}$ with $a\,R\,e$, $b\,R\,d\,R\,f$, and $c\,R\,c$. The equivalence classes are $[a]_R = \{a, e\}$, $[b]_R = \{b, d, f\}$, and $[c]_R = \{c\}$. Let $X = \{b, c\}$ then $R_*X = \{c\}$ and $R^*X = \{b, c, d, f\}$. In addition, $RN_RX = \{b, d, f\}$. The elements $\{a, e\}$ form the R–outside region of X.

Note that we always have the lower approximation of a set being contained in the upper approximation of the same set, $R_*X \subseteq R^*X$, since a subset of X surely has something (in fact everything) in common with X.

Here is a list of some of the most usefull properties of upper and lower aproximations. Upper and lower approximations are sets. To prove one set is a subset of another we show that every element of the first set is in the second set. Thus $A \subseteq B$ if we can show that an arbitrary $x \in A$ must also be in B. To prove $A = B$ we show two things; first we show $A \subseteq B$ and then we show $B \subseteq A$. Since everything in A is in B, and vice versa, A and B must be identical.

Upper and lower approximations possess the following properties:

I. $R_*X \subseteq X \subseteq R^*X$

II. $R_*\emptyset = R^*\emptyset = \emptyset$ and $R_*U = R^*U = U$

III. $R_*\left(X \cap Y\right) = R_*X \cap R_*Y$

IV. $R_*\left(X \cup Y\right) \supseteq R_*X \cup R_*Y$

V. $R^*\left(X \cup Y\right) = R^*X \cup R^*Y$

VI. $X \subseteq Y \rightarrow R_*X \subseteq R_*Y$ and $X \subseteq Y \rightarrow R^*X \subseteq R^*Y$

VII. $R_*X^{\mathsf{c}} = \left(R_*X\right)^{\mathsf{c}}$ and $R^*X^{\mathsf{c}} = \left(R^*X\right)^{\mathsf{c}}$

VIII. $R_*R_*X = R^*R_*X = R_*X$

IX. $R^*R^*X = R_*R^*X = R^*X$

As an illustration of thwew properties can be derived we now prove one of them as an example of the techniques involved. We will proove the truth of #III..

Theorem 12. $R_*\left(X \cap Y\right) = R_*X \cap R_*Y$

Proof. To prove two sets are equal you prove that each one is a subset of the other. This means that any element of the univers is in both sets or neither, so the sets must be identical. To prove that a set is a subset of another we show that every element of the first set is required to be an element of the second.

(a) We will first show that $R_*\left(X \cap Y\right) \subseteq R_*X \cap R_*Y$. Suppose that $x \in R_*\left(X \cap Y\right)$. Then by the definition of R–lower approximations, x must be in an equiavlence class $R(x)$ with $R(x) \subseteq X \cap Y$. But if $R(x)$ is contained in the intersection of A and B then it must be contained in each one individually. Thus we have that both $R(x) \subseteq A$ and $R(x) \subseteq B$. But then, again by the definition of R–lower approximation, we have that $R(x) \subseteq R_*A$

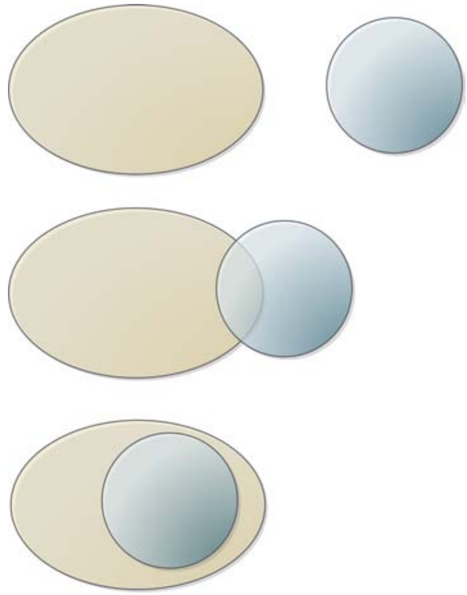


Figure 11.2.: Rough membership grade.

and $R(x) \subseteq R_*B$. Taken together, the fact that both $R(x) \subseteq R_*A$ and $R(x) \subseteq R_*B$ means that $R(x) \subseteq R_*A \cap R_*B$ and we conclude that $x \in R_*A \cap R_*B$ since, certainly, x is in $R(x)$.

(b) Next we show that $R_*(X \cap Y) \supseteq R_*X \cap R_*Y$. Suppose that $x \in R_*X \cap R_*Y$. Then by the definition of intersection x must be in both R_*A and R_*B . By the definition of R–lower approximations, x must be in an equiavlence class $R(x)$ with $R(x) \subseteq A$ and $R(x) \subseteq B$, since the equivalence class of x is unique. From the fact that $R(x) \subseteq A$ and $R(x) \subseteq B$, we conclude $R(x) \subseteq A \cap B$. If $R(x)$ is contained in the intersection of A and B then it is contained in the R–lower approximation of $A \cap B$. Since $x \in R(x)$ and $R(X) \subseteq R_*(X \cap Y)$ we have that $x \in R_*(X \cap Y)$ and we conclude that $R_*(X \cap Y) \supseteq R_*X \cap R_*Y$.

From Parts (a) and (b) we conclude that the theorem is correct. □

11.2.1. Rough Membership Function

An alternative way of defining **RST** uses a rough membership function . A rough set membership function of x in X is the ratio of the number of elements indiscernable from x that are in X to the total number of elements indiscernable from x. Thus $\mu_X^R : U \to [0,1]$ where

$$\mu_X^R(x) = \frac{|X \cap [x]_R|}{|[x]_R|} \tag{11.1}$$

The rough membership function can be thought of as the conditonal probability that x belongs to X given the information about x provided by R.

The meaning of rough membership function can be depicted as shown in Fig. 2.

While we have already defined upper an lower approximations of a set X in **RST** we now repeat this using the rough membership function can be used to define approximations and the boundary region of a set, as shown below:

$$R_*X = \left\{x \in U \mid \mu_X^R(x) = 1\right\}$$

$$R^*X = \left\{x \in U \mid \mu_X^R(x) > 0\right\}$$

$$RN_RX = \left\{x \in U \mid 0 < \mu_X^R(x) < 1\right\}$$

It is fairly easy to show that the rough membership function has the following properties.

i. $\mu_X^R(x) = 1$ iff $x \in R_*X$

ii. $\mu_X^R(x) = 0$ iff $x \in U - R^*X$

iii. $0 < \mu_X^R(x) < 1$ iff $x \in RN_RX$

iv. $\mu_{U-X}^R(x) = 1 - \mu_X^R(x)$ for any $x \in U$

v. $\mu_{X\cup Y}^R(x) \geq \max\left(\mu_X^R(x), \mu_Y^R(x)\right)$ for any $x \in U$

vi. $\mu_{X\cap Y}^R(x) \leq \min\left(\mu_X^R(x), \mu_Y^R(x)\right)$ for any $x \in U$

Example 76. Let $U = \{a, b, c, d, e, f\}$ with $a\,R\,e$, $b\,R\,d\,R\,f$, and $c\,R\,c$. The equivalence classes are $[a]_R = \{a, e\}$, $[b]_R = \{b, d, f\}$, and $[c]_R = \{c\}$. Let $X = \{b, c\}$, then the membership function $\mu_X^R(x)$ is:

x	$\mu_X^R(x)$
a	0
b	$\frac{1}{3}$
c	1
d	$\frac{1}{3}$
e	0
f	$\frac{1}{3}$

whence we $R_*X = \{c\}$ and $R^*X = \{b, c, d, f\}$. In addition, $RN_RX = \{b, d, f\}$.

From the above properties it follows that the rough membership differs fundamentally from the fuzzy membership. Properties v. and vi. indicate that the rough membership grade for x in a union or intersection of sets cannot always be computed from the membership grade of x in the individual sets. This is in stark contrast with fuzzy set theory. In general the rough membership function has characteristics that echo probability theory.

Now we can give two equivalent definitions of rough sets.

Definition 62. Set X is *rough* with respect to R if $R_*X \neq R^*X$.

Definition 63. Set X *rough* with respect to R if for some x, $0 < \mu_X^R(x) < 1$.

1

	Age	Lesions
x_1	16-30	50+
x_2	16-30	0
x_3	30-40	1-25
x_4	30-40	1-25
x_5	40-50	26-49
x_6	40-50	26-49
x_7	16-30	26-49
x_8	30-40	50+

Table 11.1.: The patents age groups and lesion counts.

	Age Group	Lesions	Beamer's Syndrome
x_1	16-30	50+	yes
x_2	16-30	0	no
x_3	30-40	1-25	no
x_4	30-40	1-25	no
x_5	40-50	26-49	no
x_6	40-50	26-49	yes
x_7	16-30	26-49	no
x_8	30-40	50+	yes

Table 11.2.: The presence of Beamer's Syndrome by age group.

Information Systems

An Information System (**IS**) is a pair (U, A) where U is a non–empty set of objects and A is set of attributes. Each individual attribute a maps U into a value set V_a, so that $a : U \to V_a$.

Example 77. Consider Table (11.1). Eight patients are classified with two attributes, Age Group and Number of Lesions observed. In the information system presented in Table (11.1) $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ and $A = \{a_1 = age, a_2 = lesion\}$. The value set of a_1 contains the age ranges of the patients: $V_{a_1} = \{16 - 30, 30 - 40, 40 - 50\}$. The value set of a_2 contains the lesion counts of the patients: $V_{a_2} = \{0, 1 - 25, 26 - 49, 50+\}$.

The same or indiscernible objects may be represented several times in an information system, for example, unbeknownst to them, two different Doctors might see the same patient, and that patients data might go into the database twice. This often happens when a name is misspelled. Some of the attributes may be superfluous, since data gatherers gather data and hope that the data turns out to be useful in the end.

Let $IS = (U, A)$ be an information system, and let B be a set of attributes, so that $B \subseteq A$. Then we say that two objects $x, y \in U$ are indiscernible with respect to B if, for ever attribute $a \in B$ the value for x is equal to the value for y, that is $\forall_{a \in B} a(x) = a(y)$.

[1]Pawlak and Skowron (1994) notes that the above definitions are not precisely equivalent.

The set of all indiscernible pairs with respect to B is called the B–indiscernability relation.

$$IND_{IS}(B) = \{\langle x, y\rangle \mid \forall a \in B,\, a(x) = a(y)\}$$

It is easy to show that indiscernability is an equivalence relation. The equivalence class of x with respect to the B—indiscernability relation is $IND_{IS}(B)$ or, more simply, $[x]_B$. The set of all equivalence classes with respect to B is denoted $U/IND_{IS}(B)$ or, more simply, U/B and

$$U/B = \{[x]_B \mid x \in U\} \ .$$

Example 78. The IS in 11.1 has $A = \{Age, Lesion\}$. With respect to Age patients x_1, x_2, and x_7 are indiscernible, that is, they are classified in the same age group. With respect to Age x_1, x_2, and x_7 are equivalent, and thus in the same equivalence class. The equivalence classes of indistinguishable Age are $\{x_1, x_2, x_7\}$, $\{x_3, x_4, x_8\}$, and $\{x_5, x_6\}$.

$$IND_{\{Age\}} = \{\{x_1, x_2, x_7\}, \{x_3, x_4, x_8\}, \{x_5, x_6\}\} \ .$$

Example 79. Let $B = \{Age, Lesion\}$ then the indiscernible groups are:

$$IND_{\{Age,Lesion\}} = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7\}, \{x_8\}\} \ .$$

11.3. Decision Systems

A decision system, **DS**, is a triple $DS = (U, C, D)$ where U is again a universal set and C and D are disjoint non–empty attribute sets. The set C is called the condition set, D is called the decision set, with $B = C \cup D$ and $C \cap D = \emptyset$.

Example 80. A **DS** with a single decision attribute d is illustrated in Table 11.2. The decision set is $D = \{Beamer\}$ and the value set of $d = Beamer$ is just yes–no; yes if the patient has been diagnosed with Beamer's Syndrome and no otherwise: $V_d = \{\text{yes}, \text{no}\}$.

All the definitions of the previous section generalize from an **IS** to a **DS** when the attribute set is disjoint collection of conditional attributes and decision attributes, $B \subseteq C \cup D$. Usually we focus on B being a subset of the decision set and hope that we learn the conditions that allow for a decision.

Example 81. The equivalence classes of $B = \{Age\}$ are $\{x_1, x_2, x_7\}$, $\{x_3, x_4, x_8\}$, and $\{x_5, x_6\}$. Let $X = \{x_4, x_5, x_6, x_7\}$ then only the equivalence class $\{x_5, x_6\}$ is contained in X. Thus the B–lower approximation of X is $\underline{B}X = \{x_5, x_6\}$. On the other hand, all of the equivalence classes induced by B contain at least one element in common with X so that the B–upper approximation of X is everything, $\bar{B}X = U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$.

Example 82. Let $B = A$ in the decision system illustrated in Table (11.2). In this case $IND_{\{Age,Lesion\}} = \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7\}, \{x_8\}\}$. Let $Y = \{x \mid Beamer(x) = \text{yes}\}$, then $Y = \{x_1, x_6, x_8\}$. Now then $\underline{A}Y = \{x_1, x_8\}$ and $\bar{A}Y = \{x_1, x_5, x_6, x_8\}$, $BN_B(Y) = \{x_1, x_5, x_6, x_8\} - \{x_1, x_8\} = \{x_5, x_6\}$ and the outside region of Y is $\{x_2, x_3, x_4, x_7\}$. We conclude that Y is rough with respect to the conditions A.

What we have shown in the examples is that *Age* and *Lesions* are not sufficient to precisely classify Beamer's Syndrome. The outside region are those patients who do not have the disease. The lower approximation are those that surely have the disease. The boundary region consists of those patients who may have the disease.

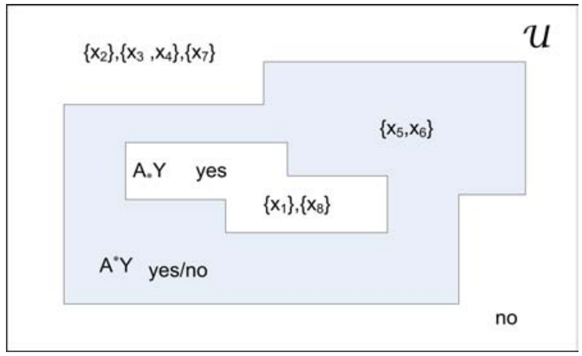


Figure 11.3.: Rough set W.

U	Cough	Temp	Bronchitis
u1	Yes	Normal	No
u2	Yes	High	Yes
u3	Yes	Very-high	Yes
u4	No	Normal	No
u5	No	High	No
u6	No	Very-high	Yes
u7	No	High	Yes
u8	No	Very-high	No

Table 11.3.: The flu.

11.4. Partitions, Covers and Rough Sets

Suppose we start with a decision system $S = (U, C, D)$. When we take a value x from the universe, such as a patient from the clinic, we derive a pair of sequences, one of condition values, and one of decision values. Suppose there are n condition attributes $\{c_1, \ldots, c_n\} = C$ and m decision attributes $\{d_1, \ldots, d_m\} = D$. Then every $x \in U$ determines a sequence of conditions $c_1(x), \ldots, c_n(x)$; and decisions $d_1(x), \ldots, d_m(x)$.

Another way to view this is to consider each sequence pair as a *decision rule* induced by x. We will denote the decision rule induced by x in S as

$$c_1(x), \ldots, c_n(x) \to d_1(x), \ldots, d_m(x) \tag{11.2}$$

or simply

$$C \to_x D\,. \tag{11.3}$$

Usually, identical decision rules are produced by many elements of the universe. We count all these elements as the support of the decision rule.

Definition 64 (support)**.** The support of a decision rule is the number of occurrences of $C \to_x D$ as u ranges over U:

$$supp_x(C, D) = |\{u \in U \mid C \to_x D = C \to_u D\}|\ . \tag{11.4}$$

Segment	In	Out	Change	Churn	N
1	medium	medium	low	no	200
2	high	high	low	no	100
3	low	low	low	no	300
4	low	low	high	yes	150
5	medium	medium	low	yes	220
6	medium	low	low	yes	30

Table 11.4.: Client segments

Mathematically, the condition and decision attribute sets are disjoint and make up all the attributes under consideration. Since $A = C \cup D$ and $\emptyset = C \cap D$ each $n+m$–tuple

$$\langle c_1(x), \ldots, c_n(x), d_1(x), \ldots, d_m(x) \rangle \tag{11.5}$$

is identical to the $n+m$–tuple

$$a(x) = \langle a_1(x), \ldots, a_{n+m}(x) \rangle \ . \tag{11.6}$$

We now provide a shorthand notation for the support by considering the equivalence class of x induced by the attributes A, traditionally $[x]_A$ but also denoted in **RST** as $A(x)$. First we note that if $[x]_C = C(x)$ and $[x]_D = D(x)$ then $A(x) = C(x) \cap D(x)$ and

$$\begin{aligned} supp_x(C, D) &= |A(x)| \\ &= |C(x) \cap D(x)| \ . \end{aligned} \tag{11.7}$$

Definition 65 (strength)**.** The strength of the decision rule induced by x is the support divided by the cardinality of U

$$\begin{aligned} \sigma_x(C, D) &= \frac{supp_x(C, D)}{|U|} \\ &= \frac{|C(x) \cap D(x)|}{|U|} \, . \end{aligned} \tag{11.8}$$

Definition 66 (certainty)**.** The certainty of the decision rule induced by x is the support divided by the cardinality of conditional equivalence class of x

$$\begin{aligned} cer_x(C, D) &= \frac{supp_x(C, D)}{|C(x)|} \\ &= \frac{|C(x) \cap D(x)|}{|C(x)|} \, . \end{aligned} \tag{11.9}$$

If $cer_x(C, D) = 1$, then $C \to_x D$ will be called a certain decision rule. If $0 < cer_x(C, D) < 1$ the decision rule will be referred to as an uncertain decision rule.

Definition 67 (coverage factor)**.** The coverage factor of the decision rule, denoted

	$H = \{1,2,3\}$	$H' = \{4,5,6\}$
$\underline{C}$	$\{2,3\}$	$\{4,6\}$
$\overline{C}$	$\{1,2,3,5\}$	$\{1,4,5,6\}$
BND	$\{1,5\}$	$\{1,5\}$

Table 11.5.: Upper an Lower Churn Approximations

Decision rule	Strength	Certainty	Coverage
1	0.20	0.48	0.33
2	0.10	1.00	0.17
3	0.30	1.00	0.50
4	0.15	1.00	0.38
5	0.22	0.52	0.55
6	0.03	1.00	0.07

Table 11.6.: Parameters of the decision rules

$cov_x(C, D)$ is defined as;

$$\begin{aligned} cov_x(C, D) &= \frac{supp_x(C, D)}{|D(x)|} \\ &= \frac{|C(x) \cap D(x)|}{|D(x)|}. \end{aligned} \tag{11.10}$$

Definition 68. Also useful is the probability of the decision equivalence classes, $\pi_x(D) = \frac{|D(x)|}{|U|}$.

It is easy to see from the formulas that the coverage factor is the ratio of the strength to the probability, or $cov_x(C, D) = \frac{\sigma_x(C,D)}{\pi_x(D)}$. It is only a slight abuse of notation (shared with probability theory) to express $cov_x(C, D) = \pi_x(C \mid D)$.

If $C \to_x D$ is a decision rule then $D \to_x C$ will be called an inverse decision rule. The inverse decision rules can be used to give explanations (reasons) for a decision.

Let us observe that if $C \to_x D$ is a decision rule then

$$\bigcup_{y \in D(x)} \{C(y) \mid C(y) \subseteq D(x)\} \tag{11.11}$$

is the lower approximation of the decision class $D(x)$, by condition classes $C(y)$, whereas the set

$$\bigcup_{y \in D(x)} \{C(y) \mid C(y) \cap D(x) \neq \emptyset\} \tag{11.12}$$

is the upper approximation of the decision class by condition classes $C(y)$.

Approximations and decision rules are two different methods to express properties of data. Approximations suit better to express topological properties of data, whereas decision rules describe in a simple way hidden patterns in data.

Example 83 (J. Grant). **Churn (from Grant (2001))**
In telecommunications, Churn refers to customers switching service providers. Table (11.4) shows a summary of 1000 clients. The attributes under consideration are:

In	incoming calls,
Out	outgoing calls within the same operator,
Change	outgoing calls to other mobile operator,
Churn	the decision attribute describing the consequence,
N	the number of similar cases.

Each row in Table (11.4) corresponds to a decision rule. For example, Row 2 determines the following decision rule: "if the number of incoming calls is high and the number of outgoing calls is high and the number of outgoing calls to the mobile operator is low then these is no churn". The N column says that 100 out 1000 clients in the database exhibit this behavior.

> One of the main problem that have to be solved by marketing departments of wireless operators is to find the way of convincing current clients that they continue to use the services. In solving this problems can help churn modeling. Churn model in telecommunications industry predicts customers who are going to leave the current operator. – J. Grant, "Churn modeling by rough set approach", manuscript, 2001.

The marketing department wants to understand why some clients Churn, Segments $\{1, 2, 3\}$, while others don't, Segments $\{4, 5, 6\}$. Note that clients in Segments 1 and 5 have identical condition attributes, but make different decisions. Thus the data is inconsistent and the problem is indeterminate.

With respect to the condition attributes, In, Out, and Change we have the following upper and lower approximations of $H = \{1, 2, 3\}$, the Churn set, and $H' = \{4, 5, 6\}$, the not–Churn set.

Segments 2 and 3 can be classified as clients who certainly do not Churn and Segments 4 and 6 can be classified as the clients who do Churn. Segments 1 and 5 are the boundary region, their behavior is undecidable.

Table (11.6) shows the strength, certainty, and coverage determined from Table (11.4).

Decision algorithm
Any decision table induces a set of "if ... then" decision rules. Any set of mutually, exclusive and exhaustive decision rules, that covers all facts in S and preserves the Indiscernability relation included by S will be called a decision algorithm in S. An example of decision algorithm derived from the decision Table (11.4) is given below:

1. if (In, high) then (Churn, no) with $cer = 1.00$
2. if (In, low) and (Change, low) then (Churn, no) with $cer = 1.00$
3. if (In, med.) and (Out, med.) then (Churn, no) with $cer = 0.48$

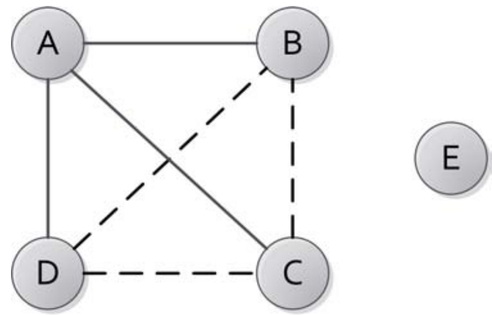


Figure 11.4.: A conflict graph.

4. if (Change, high) then (Churn, yes) with $cer = 1.00$
5. if (In, med.) and (Out, low) then (Churn, yes) with $cer = 1.00$
6. if (In, med.) and (Out, med.) then (Churn, yes) with $cer = 0.52$

Another interesting, but very complex problem, is finding a minimal decision algorithm associated for a decision table. For example, maybe we do not need to use Change to adequately explain Churn. This turns out to be a very complex problem, with many proposals as to the best method of solution.

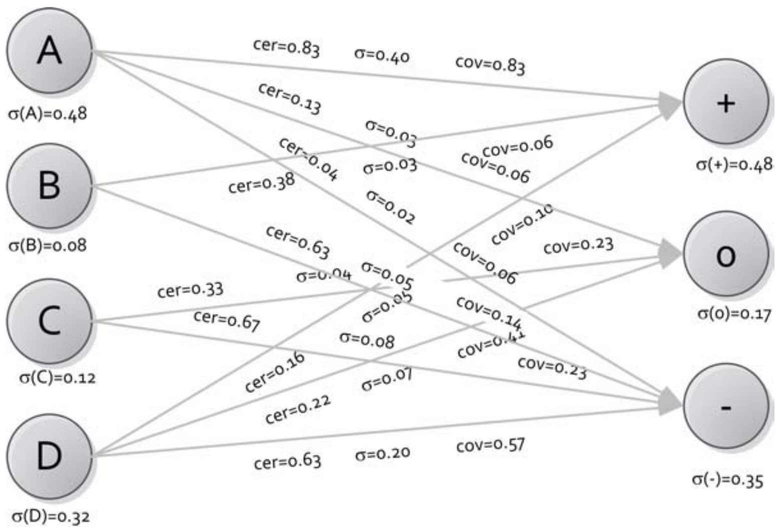
If we swap Conditions and Decision attributes we get explanatory rules. Here we list the explanatory rules (or inverse decision rules) derived from Table (11.4), along with their certainty factors.

1. if (Churn, no) then (In, high) and (Out, med.) with $cer = 0.33$
2. if (Churn, no) then (In, high) with $cer = 0.17$
3. if (Churn, no) then (In, low) and (Change, low) with $cer = 0.50$
4. if (Churn, yes) then (Change, yes) with $cer = 0.38$
5. if (Churn, yes) then (In, med.) and (Out, med.) with $cer = 0.55$
6. if (Churn, yes) then (In, med.) and (Out, low) with $cer = 0.07$

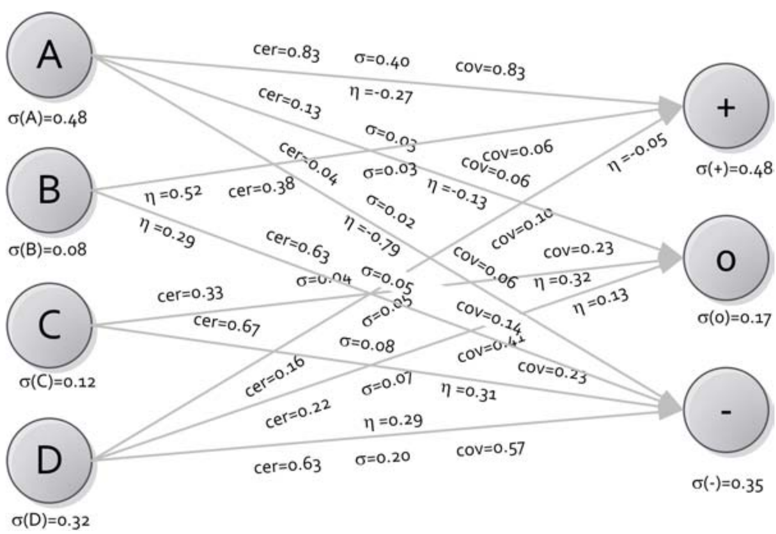
We note that that certainty factor for inverse decision rules are coverage factors for the original decision rules.

Let us summarize what the data is telling us about Churn. We conclude:

- No churn is implied with certainty by:
  - high number of incoming calls,
  - low number of incoming calls and low number of outgoing calls to other mobile operator.
- Churn is implied with certainty by:
  - high number of outgoing calls to other mobile operator,



(a) Flow graph



(b) Flow graph with dependencies.

Figure 11.5.: Flow graphs with and without dependencies.

Faction	Party	Vote	Count
1	A	+	200
2	A	0	30
3	A	-	10
4	B	+	15
5	B	-	25
6	C	0	20
7	C	-	40
8	D	+	25
9	D	0	35
10	D	-	100

Table 11.7.: A voting record

  - medium number of incoming calls and low number of outgoing calls.

- Clients with medium number of incoming calls and low number of outgoing calls within the same operator are undecided (no churn, $cer = 0.48$; churn, $cer = 0.52$).

From the inverse decision algorithm and the coverage factors we get the following explanations:

- the most probable reason for no churn is low general activity of a client,
- the most probable reason for churn is medium number of incoming calls and medium number of outgoing calls within the same operator.

11.5. Conflict Analysis

11.5.1. 1. Introduction

A.Nakamura (1999) showed how to apply **RST** concepts to conflict analysis. Conflict analysis and resolution is an important area of study in many fields such as business, political science, as well as in military operations. A student of any of these fields, political science for example, could remark that there are many different models of conflict using tools such as graph theory, Bayesian methods, differential equations, and especially game theory. The lack of success of any field , Many formal mathematical models of conflict situations have been proposed and studied using tools such as graph theory, topology, differential equations and game theory. The multitude of models shows that there is, as yet, no universal theory of conflict.

We will use a voting example to show how rough sets can aid in the analysis of conflict.

Table 11.8.: Strength, certainty and coverage factors for Table (11.7).

Fact	*Strength*	*Certainty*	*Coverage*
1	*0.4*	*0.83*	*0.83*
2	0.06	0.13	0.35
3	0.02	0.04	0.06
4	0.03	0.38	0.06
5	0.05	0.63	0.14
6	0.04	0.33	0.24
7	0.08	0.67	0.23
8	0.05	0.16	0.1
9	0.07	0.22	0.41
10	0.2	0.63	0.57

11.5.2. Basic concepts of conflict theory

Let U be a finite, non-empty set whose members will be referred to as agents. The attribute function v (for vote) maps agents to the values:

-1 representing *against*,

0 representing abstain, and

+1 representing *for*.

Thus $v : U \to \{-1, 0, 1\}$, or in short $v : U \to \{-, 0, +\}$. The pair $S = (U, v)$ will be called a conflict situation. In order to express relations between agents we define three basic binary relations between agents: conflict, neutrality and alliance. To this end we first define the following auxiliary function:

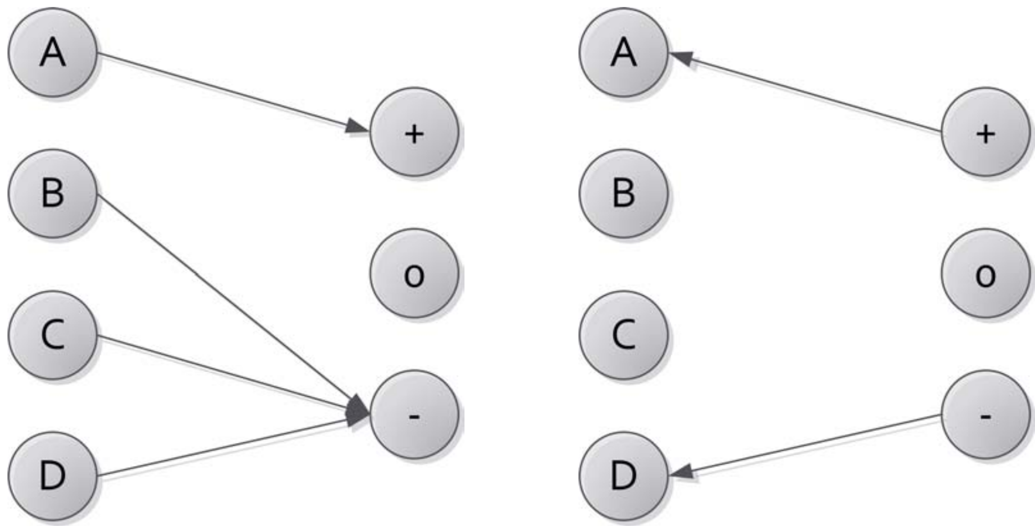
$$\phi_v(x,y) = \begin{cases} 1 & x = y \\ 0 & v(x)v(y) = 0 \,\text{and}\, x \neq y \\ -1 & v(x)v(y) = -1 \end{cases} \tag{11.13}$$

We interpret ϕ_v thusly:

$\phi_v(x,y) = 1$ this means means agents x and y have the same opinion about issue v (are allied on v),

$\phi_v(x,y) = 0$ means that at exactly one agent x or y has a neutral approach to the issue, and

$\phi_v(x,y) = -1$ means that both agents have different opinions about issue v (are in conflict on v).



(a) Forward Approximate Flow Graph (b) Inverse Approximate Flow Graph

Figure 11.6.: Approximate flow graphs.

We now define three basic relations on $U \times U$ called alliance, R_v^+, neutrality , R_v^0, and conflict, R_v^-, relations respectively. The definitions of alliance, neutrality, and conflict relations are as follows:

$$R_v^+(x,y) \text{ iff } \phi_v(x,y) = 1$$

$$R_v^0(x,y) \text{ iff } \phi_v(x,y) = 0$$

$$R_v^-(x,y) \text{ iff } \phi_v(x,y) = -1$$

Since $R_v^+(x,y) \text{ iff } \phi_v(x,y) = 1$ and $\phi_v(x,y) = 1 \text{ iff } x = y$ it is obvious that alliance, $R_v^+(x,y)$ is an equivalence relation, and thus verifies the following properties:

(i) $R_v^+(x,x)$ for all $x \in U$

(ii) $R_v^+(x,y)$ implies $R_v^+(y,x)$

(iii) $R_v^+(x,y)$ and $R_v^+(y,z)$ imply $R_v^+(x,z)$.

The equivalence classes of alliance are called coalitions with respect to v. We see that item ((iii)) says that "the friend of my friend is a friend."

Since $R_v^-(x,y) \text{ iff } \phi_v(x,y) = -1$ and $\phi_v(x,y) = -1 \text{ iff } v(x)v(y) = -1$ it is obvious that conflict, $R_v^+(x,y)$ is not an equivalence relation. However it does possess the following properties:

(iv) not $R_v^-(x,x)$ for all $x \in U$

(v) $R_v^0(x,y)$ implies $R_v^0(y,x)$

(vi) $R_v^-(x,y)$ and $R_v^-(y,z)$ imply $R_v^+(x,z)$

(vii) $R_v^-(x,y)$ and $R_v^+(y,z)$ imply $R_v^-(x,z)$.

We see that item ((vii)) says that "the enemy of my freind is my enemy."

Since $R_v^0(x,y)\,\mathrm{iff}\,\phi_v(x,y) = 0$ and $\phi_v(x,y) = o\,\mathrm{iff}\,\mathrm{v(x)v(y)} = 0\,\mathrm{and}\,x \neq y$ it is obvious that neutrality, $R_v^)(x,y)$ verifies the following properties:

(viii) not $R_v^0(x,x)$ for all $x \in U$

(ix) $R_v^0(x,y)$ implies $R_v^0(y,x)$

(x) $R_v^+(x,y)$ and $R_v^+(y,z)$ imply $R_v^+(x,z)$.

Let us observe that in the conflict and neutrality relations there are no coalitions. The following property holds:

$$R_v^+(x,y) \cup R_v^0(x,y) \cup R_v^-(x,y) = U^2$$

because one of the three cases for ϕ_v must hold, that is, given an x and y one of the three case in Eq. (11.13) must hold. If the values of $v(x)$ and $v(y)$ are equal, then the first case, $\phi_v(x,y) = 1$ is true and $R_v^+(x,y)$ holds so x and y agree. If $v(x)$ is not equal to $v(y)$ then if exactly one is zero then the second case of Eq. (11.13) is true and $R_v^0(x,y)$ holds so x and y are neutral to each other. Otherwise there is conflict as one of the values of $v(x)$ and $v(y)$ are plus one and minus one, or the reverse, and the product is certainly negative one so $v(x)v(y) = -1$ and $R_v^-(x,y)$ holds.

With every conflict situation we will associate a conflict graph G.

An example of a conflict graph is shown in Fig. 11.4. In Fig. 11.4 solid lines are denoting conflicts, doted line − alliance, and neutrality, for simplicity, is not shown explicitly in the graph. Of course, B, C, and D form a coalition.

11.6. An example

In this section we will illustrate the above presented ideas by means of a very simple tutorial example using concepts presented in the previous. Table 1 presents a decision table in which the only condition attribute is Party, whereas the decision attribute is Voting. The table describes voting results in a parliament containing 500 members grouped in four political parties denoted A, B, C and D. Suppose the parliament discussed certain issue (e.g., membership of the country in European Union) and the voting result is presented in column Voting, where +, 0 and − denoted yes, abstention and no respectively. The column support contains the number of voters for each option.

The strength, certainty and the coverage factors for Table 11.7 are given in Table 11.8.

From the certainty factors we can conclude, for example, that:

- 83.3% of party A voted *yes*,
- 12.5% of party A *abstained*,
- 4.2% of party A voted *no*.

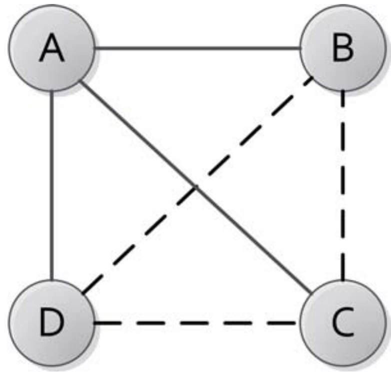


Figure 11.7.: Conflict graph associated with Figure 11.6.

From the coverage factors we can get, for example, the following explanation of voting:

- 83.3% *yes* votes came from party A
- 6.3% *yes* votes came from party B
- 10.4% *yes* votes came from party C

The flow graph associated with Table 2 is shown in Fig. 2.

Branches of the flow graph represent decision rules together with their certainty and coverage factors. For example, the decision rule $A \to 0$ has the certainty and coverage factors 0.125 and 0.353, respectively. The flow graph gives a clear insight into the voting structure of all parties.

For many applications exact values of certainty of coverage factors of decision rules are not necessary. To this end we introduce “approximate” decision rules, denoted $C \Rightarrow D$ and read “C mostly implies D”. $C \Rightarrow D$ if and only if $cer(C, D) > 0.5$. Thus we can replace flow graph shown in Fig. 11.5a by "approximate” flow graph presented in Fig. 11.6a.

Fig. 11.6b is a flow graph contains all inverse decision rules with certainty factor greater than 0.5. From this graph we can see that *yes* votes were obtained mostly from party A and no votes − mostly from party D.

From the graph 11.6a we can see that parties B, C and D form a coalition, which is in conflict with party A, i.e., every member of the coalition is in conflict with party A. The corresponding conflict graph is shown in Fig. 11.4.

A useful metric for analysis of flow graphs is the dependency factor, $\eta(x, y)$, given by:

$$\eta(x, y) = \frac{cer(x, y) - \sigma(y)}{cer(x, y) + \sigma(y)} .$$

Fig. 11.5b shows the dependency factor for the voting example.

Homework

Given the data :

	Height	Weight	Color
x_1	5	200+	red
x_2	5	100-150	red
x_3	5	150-200	blue
x_4	6	100-150	blue
x_5	6	150-200	red
x_6	5	100-150	blue
x_7	5	200+	blue
x_8	6	150-200	red

1. What is $IND_{\{Height\}}$?
2. What is $IND_{\{Weight\}}$?
3. What is $IND_{\{Height,Weight\}}$?
4. What is $IND_{\{Color\}}$?
5. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$. Let $W = \{x5, x6\}$. What is $\underline{D}W$?
6. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$.Let $W = \{x5, x6\}$. What is $\overline{D}W$?
7. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$.Let $Y = \{x \mid Height(x) = 4\}$. What is $\underline{D}Y$?
8. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$.Let $Y = \{x \mid Height(x) = 4\}$. What is $\overline{D}Y$?
9. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$.Let $Z = \{x \mid Weight(x) = 200+\}$. What is $\underline{D}Z$?
10. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$.Let $Z = \{x \mid Weight(x) = 200+\}$. What is $\overline{D}Z$?
11. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$. What is $supp_x(C, D)$, $\sigma_x(C, D)$, $cer_x(C, D)$ and $cov_x(C, D)$ when $x = x_1$?
12. Suppose the data represents a decision system with $C = \{Height, Weight\}$ and $D = \{color\}$. What is $supp_x(C, D)$, $\sigma_x(C, D)$, $cer_x(C, D)$ and $cov_x(C, D)$ when $x = x_2$?

Suppose we have the following decision system (where N represents a frequency):

	Length	Years	Disease	N
1	3	20	yes	50
2	5	10	yes	100
3	5	15	yes	200
4	4	10	no	100
5	3	15	yes	200
6	4	20	no	100

13. Fill in the following table based on the above data:

Decision rule	Strength	Certainty	Coverage
1			
2			
3			
4			
5			
6			

Suppose we have the following Voting Data

Faction	Party	Vote	Count
1	A	+	200
2	A	0	30
3	A	-	60
4	B	+	75
5	B	-	25
6	B	0	20
7	C	0	40
8	C	-	60

14. Give the above voting data, fill out the following table:

Faction	*Strength*	*Certainty*	*Coverage*
1			
2			
3			
4			
5			
6			
7			
8			

15. Give the above voting data, create a flow graph.

12. Evidence Theory

12.1. Introduction

In classical mathematics measures are additive. If Ann brings in 5 pounds of apples and Bob brings in 6 pounds of bananas we have $5 + 6 = 11$ pounds of fruit.

Suppose on the other hand that Ann and Bob are witnesses to a crime. Take individually, their testimony would convince a juror that there is a 50% likelihood that Carl is guilty. Yet taken together, their testimony will not convince a juror that Carl is guilty without a doubt. That is because evidence, in general, is non–additive.

Th mathematics of non–additive measures is briefly explained in Section A.2 on page 314 of the Appendices. For our purposes, the important characteristics of fuzzy measures are that it assigns numbers to sets and that it is monotone. A fuzzy measure g is monotone if $A \subseteq B$ implies that $g(A) \leq g(B)$.

12.2. Evidence Theory

Evidence theory (**ET**) is one of the broadest frameworks for the representation of uncertainty. Its origins lie in the works of Dempster (1967b,a) and Shafer (1976) are heavily influenced by probability theory, one of the oldest uncertainty frameworks. Evidence theory is especially important because it is a kind of Swiss army knife in the field of uncertainty. **ET** encompasses belief, plausibility, necessity, possibility and probability among a host of other measures. Here we present Evidence Theory as it was originally characterized by Shafer.

Evidence theory is based on two fuzzy measures: belief measures and plausibility measures. Belief and plausibility measures can be conveniently characterized by a function m from the power set of the universal set X into the unit interval. In this chapter we will assume at all times that X is finite. The function m, where $m : \mathcal{P}(X) \to [0, 1]$, is required to satisfy two conditions:

$$\begin{aligned} &(1) \qquad m(\emptyset) = 0 \\ &(2) \qquad \sum_{A \in \mathcal{P}(X)} m(A) = 1. \end{aligned} \tag{12.1}$$

The function m is called a *basic probability assignment* (**bpa**). For each set $A \in \mathcal{P}(X)$, the value $m(A)$ expresses the proportion to which all available and relevant evidence supports the claim that a particular element of X belongs to the set A. This value, $m(A)$, pertains solely to one set, set A; it does not imply any additional claims regarding subsets or supersets of A. If there is some additional evidence supporting the claim that the element belongs to a subset of A, say $B \subseteq A$, it must be expressed by another value $m(B)$.

Given a basic probability assignment, m, every set $A \in \mathcal{P}(X)$ for which $m(A) \neq 0$ is called a *focal element*. The pair $\langle \mathcal{F}, m \rangle$, where $\mathcal{F}$ denotes the set of all focal elements induced by m is called a *body of evidence* and we may denote it by $\mathcal{B} = \langle \mathcal{F}, m \rangle$.

From a basic probability assignment m, the corresponding belief measure and plausibility measure are determined for all sets $A \in \mathcal{P}(X)$ by the formulas

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B), \tag{12.2}$$

and

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{12.3}$$

Thus the belief in a set A is the sum of all the evidence (basic probability) assigned to A or to any subset of A. On the other hand, the plausibility of A is the sum of all the evidence (basic probability) that overlaps with A.

It can be shown that the plausibility of an event is one minus the belief of the complement of that event, and vice verse. That is

$$\mathrm{Bel}(A) = 1 - \mathrm{Pl}(A^{\mathsf{c}})$$
$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^{\mathsf{c}})$$

Since we can calculate the belief from the plausibility, and the plausibility from the belief, and both belief and plausibility can be derived from the basic probability assignment, we only need one more formula to show that all three measures provide the same information.

Given a belief measure Bel, the corresponding basic probability assignment m is determined for all $A \in \mathcal{P}(X)$ by the formula

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B), \tag{12.4}$$

where $|A - B|$ is the cardinality of the set difference of A and B, as proven by Shafer (1976). [1] Thus each of the three function, m, Bel and Pl, is sufficient to determine the other two.

Total ignorance is expressed in evidence theory by $m(X) = 1$ and $m(A) = 0$ for all $A \neq X$. Full certainty is expressed by $m(\{x\}) = 1$ for one particular element of x and $m(A) = 0$ for all $A \neq \{x\}$.

Example 84. As an example, let $X = \{x_1, x_2, x_3\}$ and let

[1] This book uses $A \setminus B$ for set difference, but this is the traditional presentation of the formulas of evidence theory in most books and papers.

$$\begin{aligned} m(\{x_1, x_2\}) &= 0.3 \\ m(\{x_3\}) &= 0.1 \\ m(\{x_2, x_3\}) &= 0.2 \\ m(X) &= 0.4\,. \end{aligned} \tag{12.5}$$

be a given basic probability assignment on $\mathcal{P}(X)$. The focal set of this basic probability assignment is the set

$$\mathcal{F} = \{\{x_1, x_2\}, \{x_3\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}\}; \tag{12.6}$$

and we always assume that $m(A) = 0$ for all $A \notin \mathcal{F}$, that is, m is zero for any sets that are not listed or mentioned.

Using the given basic probability assignment we can calculate the belief and plausibility of any subset of X. For example, our belief in $\{x_2, x_3\}$ is

$$\begin{aligned} \mathrm{Bel}(\{x_2, x_3\}) &= m(\{x_2, x_3\}) + m(\{x_3\}) \\ &= 0.2 + 0.1 \\ &= 0.3. \end{aligned} \tag{12.7}$$

since $\{x_2, x_3\}$ and $\{x_3\}$ are the only subsets of $\{x_2, x_3\}$ in the focal set. The plausibility of $\{x_3\}$ is

$$\begin{aligned} \mathrm{Pl}(\{x_3\}) &= m(X) + m(\{x_2, x_3\}) + m(\{x_3\}) \\ &= 0.4 + 0.2 + 0.1 \\ &= 0.7, \end{aligned} \tag{12.8}$$

since X, $\{x_2, x_3\}$, and $\{x_3\}$ are in the focal set and their intersection with $\{x_3\}$ is non-empty.

Table 12.1 is complete listing of the basic probability assignment, belief, and plausibility of all subsets of X for this example.

Two special cases of evidence are important.

12.2.1. Probability Theory

In the first special case, suppose that each element of the focal set has size one. Thus each set $A_i \in \mathcal{F}$ contains a single element of the universe X. In this case, since the sum of the weights of the focal elements is one, and each focal element element contains one object from the universe. Define $p : X \to [0, 1]$ by

$$p(x) = m(\{x\})$$

then p is always positive, and sums to one. Since we assume X is finite p is a probability distribution, and the belief, plausibility, and possibility of a subset A of X are

Table 12.1.: An example of a basic probability assignment and the associated belief and plausibility measures.

Set	m	Bel	Pl
$\emptyset$	0.0	0.0	0.0
$\{x_1\}$	0.0	0.0	0.7
$\{x_2\}$	0.0	0.0	0.9
$\{x_1, x_2\}$	0.3	0.3	0.9
$\{x_3\}$	0.1	0.1	0.7
$\{x_1, x_3\}$	0.0	0.1	1.0
$\{x_2, x_3\}$	0.2	0.3	1.0
X	0.4	1.0	1.0

all identical, $Bel(A) = Pl(A) = P(A)$.

Example 85. Let us examine the following body of evidence defined on $X = \{x_1, x_2, x_3\}$.

$$\begin{aligned} m(\{x_1\}) &= 0.3 \\ m(\{x_2\}) &= 0.2 \\ m(\{x_3\}) &= 0.5 \end{aligned}$$

The focal set is $\mathcal{F} = \{\{x_1\}, \{x_2\}, \{x_3\}\}$. We note that each focal object contains a single element. Define

$$\begin{aligned} p_1 &= p(x_1) = m(\{x_1\}) = 0.3 \\ p_2 &= p(x_2) = m(\{x_2\}) = 0.2 \\ p_3 &= p(x_3) = m(\{x_3\}) = 0.5 \end{aligned}$$

and we claim that p is a probability distribution on X. It satisfies all the requirement of a discrete probability distribution, all the probabilities are non–negative and the probabilities sum to one. The following Table shows the **bpa**, belief, plausibility, and probability of all the subsets of X.

Set	m	Bel	Pl	P
$\emptyset$	0	0	0	0
$\{x_1\}$	0.3	0.3	0.3	0.3
$\{x_2\}$	0.2	0.2	0.2	0.2
$\{x_1, x_2\}$	0	0.5	0.5	0.5
$\{x_3\}$	0.5	0.5	0.5	0.5
$\{x_1, x_3\}$	0	0.8	0.8	0.8
$\{x_2, x_3\}$	0	0.7	0.7	0.7
X	0	1	1	1

12.2.2. Possibility Theory

In the second special case, assume that the elements of the focal set are consonant. This means that there is an order of the focal sets such that each set is nested in its successors. That is, if A_i and A_j are elements of the focal set $\mathcal{F}$ then, if $i < j$ then $A_i \subset A_j$.

The special branch of evidence theory that deals only with bodies of evidence whose focal elements are nested is referred to as *possibility theory* [Dubois and Prade, 1988]. Special counterparts of belief measures and plausibility measures in possibility theory are called *necessity measures* and *possibility measures*, respectively. Thus $Pos(A) = Pl(A)$ and $Nec(A) = Bel(A)$

It is not hard to show that in the case of consonant focal sets that:

i. $Nec(A \cap B) = \min[Nec(A), \mathrm{Nec}(B)]$ for all $A, B \in \mathcal{P}(X)$;

ii. $Pos(A \cup B) = \max[\mathrm{Pos}(A), \mathrm{Pos}(B)]$ for all $A, B \in \mathcal{P}(X)$.

Define

$$r(x) = \max_{x \in A} [Pos(A)]$$

then r is called a possibility distribution. We now that the values of r must be between 0 and 1 so that $r{:}X \to [0,1]$. It also turns out that

$$Pos(A) = \max_{x \in A} r(x) \tag{12.9}$$

for each $A \in \mathcal{P}(X)$.

Example 86. Let us examine the following body of evidence defined on $X = \{x_1, x_2, x_3\}$.

$$\begin{aligned} m(\{x_1\}) &= 0.3 \\ m(\{x_1, x_2\}) &= 0.2 \\ m(\{x_1, x_2, x_3\}) &= 0.5 \end{aligned}$$

The focal set is $\mathcal{F} = \{\{x_1\}, \{x_1, x_2\}, \{x_1, x_2, x_3\}\}$. We note that the focal object can be ordered by containment, $\{x_1\} \subseteq \{x_1, x_2\} \subseteq \{x_1, x_2, x_3\}$. The following Table shows the **bpa**, possibility, and necessity of all the subsets of X.

Set	m	Pos	Nec
$\emptyset$	0	0	0
$\{x_1\}$	0.3	0.3	1
$\{x_2\}$	0	0.0	0.7
$\{x_1, x_2\}$	0.2	0.5	1
$\{x_3\}$	0	0	0.5
$\{x_1, x_3\}$	0	0.3	1
$\{x_2, x_3\}$	0	0	0.7
X	0.5	1	1

Define a possibility distribution on X

$$\begin{aligned} r_1 &= r(x_1) = & \max_{x_1 \in A} [Pos(A)] = 0.3 \\ r_2 &= r(x_2) = & \max_{x_1 \in A} [Pos(A)] = 0.5 \\ r_3 &= r(x_3) = & \max_{x_1 \in A} [Pos(A)] = 1 \end{aligned}$$

then for each subset A of X we have that the possibility of A is just the largest distributional value assigned to an element of A.

12.2.2.1. Possibility Theory as Fuzzy Set Theory

The most visible interpretation of possibility theory, is connected with fuzzy sets. This interpretation was introduced by Zadeh [1978].

To explain the fuzzy set interpretation of possibility theory, let $\mathcal{X}$ denote a variable that takes values in a given set X, and let information about the actual value of the variable be expressed by a fuzzy proposition "$\mathcal{X}$ *is* F", where F is a standard fuzzy subset of X (i.e., $F(x) \in [0,1]$ for all $x \in X$). To express information in measure-theoretic terms, it is natural to interpret the membership degree $F(x)$ for each $x \in X$ as the degree of possibility that $\mathcal{X} = x$. This interpretation induces a possibility distribution, r_F, on X that is defined by the equation

$$r_F(x) = F(x) \tag{12.10}$$

for all $x \in X$. If F is normal then a fuzzy set is, in a sense, equivalent to a possibility distribution.

Example 87. Let F be the normal fuzzy set $X = \{x_1, x_2, x_3\}$ defined by

$$\begin{aligned} F(x_1) &= 0.3 \\ F(x_1) &= 0.5 \\ F(x_1) &= 1.0 \end{aligned}$$

Define a possibility distribution r_F on X

$$\begin{aligned} r_F(x_1) &= 0.3 \\ r_F(x_2) &= 0.5 \\ r_F(x_3) &= 1\,. \end{aligned}$$

We note that the normal fuzzy set F, the possibility distribution r_F and the possibility distribution of Example 86 are all identical.

Homework

Let $X = \{x_1, x_2, x_3\}$ and let

$m(\{x_1\})$	$= 0.1$
$m(\{x_3\})$	$= 0.3$
$m(\{x_1, x_2\})$	$= 0.4$
$m(X)$	$= 0.2$
Body of Evidence 1	

$m(\{x_1, x_3\})$	$= 0.2$
$m(\{x_2, x_3\})$	$= 0.3$
$m(\{x_1, x_2\})$	$= 0.4$
$m(X)$	$= 0.1$
Body of Evidence 2	

$m(\{x_1\})$	$= 0.1$
$m(\{x_1, x_2\})$	$= 0.3$
$m(X)$	$= 0.6$
Body of Evidence 3	

$m(\{x_3\})$	$= 0.1$
$m(\{x_2, x_3\})$	$= 0.4$
$m(X)$	$= 0.2$
Body of Evidence 4	

$m(\{x_1\})$	$= 0.1$
$m(\{x_2\})$	$= 0.3$
$m(\{x_3\})$	$= 0.6$
Body of Evidence 5	

$m(\{x_1\})$	$= 0.7$
$m(\{x_2\})$	$= 0.3$
Body of Evidence 6	

1. Given the **bpa** in *Body of Evidence 1* what are the focal elements.

2. Given the **bpa** in *Body of Evidence 1* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

3. Given the **bpa** in *Body of Evidence 2* what are the focal elements.

4. Given the **bpa** in *Body of Evidence 2* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

5. Given the **bpa** in *Body of Evidence 3* what are the focal elements.

6. Given the **bpa** in *Body of Evidence 3* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

7. Given the **bpa** in *Body of Evidence 4* what are the focal elements.

8. Given the **bpa** in *Body of Evidence 4* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

9. Given the **bpa** in *Body of Evidence 5* what are the focal elements.
10. Given the **bpa** in *Body of Evidence 5* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

11. Given the **bpa** in *Body of Evidence 6* what are the focal elements.
12. Given the **bpa** in *Body of Evidence 6* fill in the following Table:

Set	m	Bel	Pl
$\emptyset$			
$\{x_1\}$			
$\{x_2\}$			
$\{x_1, x_2\}$			
$\{x_3\}$			
$\{x_1, x_3\}$			
$\{x_2, x_3\}$			
X			

13. Are any of *Body of Evidence 1–Body of Evidence 6* probability distributions? How about possibility distributions?
14. Fill in the following Table:

Set	m	Bel	Pl
$\emptyset$		0.0	
$\{x_1\}$		0.0	
$\{x_2\}$		0.0	
$\{x_1, x_2\}$		0.4	
$\{x_3\}$		0.1	
$\{x_1, x_3\}$		0.1	
$\{x_2, x_3\}$		0.4	
X		1.0	

13. Data Fusion

13.1. Dempster's Rule

Suppose we have two bodies of evidence, one from Ann and one from Bob. Suppose we want to fuse this data into a single body of evidence. We assume that both bodies of evidence concern the same universe X bt that Ann and Bob view the situation differently. Let Ann's evidence be $B_1 = \langle \mathcal{F}_1, m_1 \rangle$ and Bob's evidence be $B_2 = \langle \mathcal{F}_2, m_2 \rangle$. Let K be the conflict among the bodies evidence. Conflict occurs when Ann and Bob focus evidence on focal sets that have nothing in common. Ann's evidence for $A \in \mathcal{F}_1$ and Bob's evidence for $B \in \mathcal{F}_2$ conflict whenever $A \cap B = \emptyset$. Defince the total conflict as

$$\mathcal{K} = \sum_{A \cap B = \emptyset} m_1(A)\, m_2(B)$$

where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Then we can create a fused body of evidence $B = \langle \mathcal{F}, m \rangle$ from Ann's evidence $B_1 = \langle \mathcal{F}_1, m_1 \rangle$ and Bob's evidence be $B_2 = \langle \mathcal{F}_2, m_2 \rangle$ by setting the focal set of $B = \langle \mathcal{F}, m \rangle$ to

$$\mathcal{F} = \{A \cap B \mid A \cap B \neq \emptyset \text{ and } A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$$

and defining the basic probability assignment of $\mathcal{B} = \langle \mathcal{F}, m \rangle$ for $C \in \mathcal{F}$ as

$$m(C) = \frac{\sum_{A \cap B = C} m_1(A)\, m_2(B)}{1 - \mathcal{K}}$$

If we do not divide by the conflict term $1 - \mathcal{K}$ then m will not sum to one. The normalization factor above, $1 - \mathcal{K}$, has the effect of completely ignoring conflict and attributing any mass associated with conflict to the null set. This combination rule for evidence can therefore produce counterintuitive results when there is significant conflict or not.

Example 88. Let us examine the following bodies of evidence defined on three suspects of a crime $x_1 = Ralph$, $x_2 = Sue$, and $x_3 = Tom$, so that $X = \{x_1, x_2, x_3\}$.

Ann is an invistigative reporter and determins that there are four critical pieces of evidence, that assign guilt thusly,

$$\begin{aligned} m_1(\{x_1, x_2\}) &= 0.3 \\ m_1(\{x_3\}) &= 0.1 \\ m_1(\{x_2, x_3\}) &= 0.2 \\ m_1(X) &= 0.4\,. \end{aligned} \tag{13.1}$$

We interpret $m(\{x_1, x_2\}) = 0.3$ as one piece of evidence, say eye–witness testimony, that places Ralph and Sue in proximity of the crime scene. The reliability of the eye–witness causes us to weight this evidence as having 30% of the available credibility. The focal set for this body of evidence is:

$$\mathcal{F}_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}\}; \tag{13.2}$$

Bob is NCIS and arrives at the follwing body of evidence as a result of his investigation.

$$\begin{aligned} m_2(\{x_1\}) &= 0.3 \\ m_2(\{x_1, x_2\}) &= 0.2 \\ m_2(\{x_1, x_2, x_3\}) &= 0.5 \end{aligned} \tag{13.3}$$

The focal set for this body of evidence is

$$\mathcal{F}_2 = \{\{x_1\}, \{x_1, x_2\}, \{x_1, x_2, x_3\}\} . \tag{13.4}$$

Conflict between bodies of evidence ocurs when focal sets have an empty intersection. For example $\{x_3\} \cap \{x_1\} = \emptyset$ so the product of their weights $m_1(\{x_3\})$ and $m_2(\{x_1\})$, or 0.1×0.3, gets added to $\mathcal{K}$.

$$\begin{aligned} \mathcal{K} &= m_1(\{x_3\}) \times m_2(\{x_1\}) + m_1(\{x_3\}_1) \times m_2(\{x_1, x_2\}) + m_1(\{x_2, x_3\}_1) \times m_2(\{x_1\}) \\ &= 0.1 \times 0.3 + 0.1 \times 0.2 + 0.2 \times 0.3 \\ &= 0.13 \end{aligned}$$

In the fused body of evidence $m_{12}(\{x_1\})$ would be the sum of all the evidence where Ann and Bob's evidence agrees only on $\{x_1\}$ divided by $1 - \mathcal{K}$. Thus:

$$\begin{aligned} m_{12}(\{x_1\}) &= \frac{m_1(\{x_1, x_2\}) \times m_2(\{x_1\}) + m_1(X) \times m_2(\{x_1\})}{1 - \mathcal{K}} \\ &= \frac{0.3 \times 0.3 + 0.4 \times 0.3}{0.87} \\ &= 0.2414 \end{aligned}$$

The following Table shows the **bpa**, possibility, and necessity of the fused body of evidence for all the subsets of X.

Set	m	Bel	Pl
$\emptyset$	0.00	0.00	0.00
$\{x_1\}$	0.43	0.43	0.43
$\{x_2\}$	0.20	0.20	0.20
$\{x_1, x_2\}$	0.00	0.64	0.64
$\{x_3\}$	0.36	0.36	0.36
$\{x_1, x_3\}$	0.00	0.80	0.80
$\{x_2, x_3\}$	0.00	0.57	0.57
X	0.00	1.00	1.00

Define a possibility distribution on X

Zadeh is also important to evidence theory due to his critique of Dempster's rule of combination Zadeh (1984). The following example is due to Zadeh.

Example 89 (Effect of high conflict.)**.** Suppose that one doctor believes a patient has either meningitis — with a probability of 0.99 — or a brain tumor — with a probability of only 0.01. A second doctor believes the patient suffers from concussion — with a probability of 0.99 — and also believes the patient has a brain tumor — with a probability of only 0.01. Since the only non–conflicting diagnosis is brain tumor, it is the only result with a positive basic probability. When we normalize it, we will get that $m(brain\,tumor) = 1$ and hence $Bel(brain\,tumor) = 1$ so that we have total beleif in a brain tumor even though each doctor beleives this diagnosis is a remote possibility.

Example 90 (Effect of low conflict.)**.** Suppose that one doctor believes a patient has either a brain tumor — with a probability of 0.99 — or meningitis — with a probability of only 0.01. A second doctor also believes the patient has a brain tumor — with a probability of 0.99 — and believes the patient suffers from concussion — with a probability of only 0.01. Since the only non–conflicting diagnosis is brain tumor, it is the only result with a positive basic probability. When we normalize it, we will get that $m(brain\,tumor) = 1$ and hence $Bel(brain\,tumor) = 1$ so that we have total beleif in a brain tumor even though neither doctor is absolutely sure of his diagnosis.

In both of the above examples we the same result, which is counter to our intuition. As a result, many alternative methods of fusion have been proposed. One easy way out of this delemma is not to normalize the results. OblowOblow (1986) was one of the first to propose that evidence be allowed to accumulate on the empty set in a variation of evidence theory he termed O-Theory. Yager Yager (1987) was the first to propose this fix for data fusion. Many modern applications of evidence theory assume that $m(\emptyset)$ can be positive, especially in data fusion.

13.2. Reaching a Verdict by Weighting Evidence

We will now examine some alternative methods of fusing evidence based on weighting the experts. We will see how the classical decision making rules formulated by Hooper, Dempster, Bayes, and Jeffrey are special cases of weighting bodies of evidence. The case of a body of evidence induced by a fuzzy set is also introduced.

A body of evidence induces a probability (or credibility) distribution on the class $\mathcal{P}(X)$ of all possible subsets of X. We will assume that evidence can now reside upon the empty set.

Thus $m_i\colon P(X) \longrightarrow [0,1]$ and

$$\begin{aligned} &(1) \qquad m(\emptyset) \geq 0 \\ &(2) \qquad \sum_{A \in \mathcal{P}(X)} m_i(A) = 1. \end{aligned} \tag{13.5}$$

The number $m_i(A)$ denotes the probability (or credibility) that the suspects belong to the subset A but not to a subset of it. The class of $focal$ subsets of X corresponding to m_i is

$$F_i = \{A \mid A \subseteq X,\, m_i(A) > 0\} \ . \tag{13.6}$$

The belief, plausibility and ambiguity of A induced by m_i are defined as

$$Bel_i(A) = \sum_{B \subseteq A,\, B \neq \emptyset} m_i(B),$$

$$Pl_i(A) = \sum_{B \cap A \neq \emptyset} m_i(B) \ .$$

$$Amb_i(A) = \sum_{B \cap A \neq \emptyset, B \nsubseteq A} m_i(B) \, .$$

Definition 69 (Simple evidence)**.** Simple evidence refers to the case when the bodies of evidence are mutually independent. A body of evidence induces a probability (credibility) distribution on $\mathcal{P}(X)$. Thus $m_i \colon \mathcal{P}(X) \longrightarrow [0,1]$ is a **bpa**. The number $m_i(A)$ denotes the probability (or credibility) that the suspects belong to the subset A but not to a subset of it.

Definition 70 (Mixed evidence)**.** A pair of dependent bodies of evidence, let us say witness i and witness j testifying dependently, induce a joint probability (credibility) distribution, namely $m_{ij} : P(X) \times P(X) \longrightarrow [0,1]$, where $m_{ij}(A,B)$ is the probability (credibility) that witness i focuses on subset A and witness j focuses on subset B. If the bodies of evidence are independent, then $m_{ij}(A,B) = m_i(A)\, m_j(B)$. If $m_j(B) > 0$, the conditional probability (credibility) distribution on $P(X)$ given B is $m_{i|j}(A \mid B) = m_{ij}(A,B) / m_j(B)$. The corresponding class of focal pairs of subsets is

$$F_{ij} = \{(A,B) \mid A \subseteq X,\, B \subseteq X,\, m_{ij}(A,B) > 0\}$$

.

In a natural way we can introduce the functions $BelBel$, $BelPl$, $PlPl$, $BelAmb$, etc., on $P(X)$. Thus, for instance,

$$BelPl_{ij}(A,B) = \sum_{C \subseteq A,\, C \neq \emptyset} \sum_{D \cap B \neq \emptyset} m_{ij}(C,D) \, .$$

Obviously, if the bodies of evidence i and j are independent, then $BelPl_{ij}$ is equal to $Bel_i \times Pl_j$, or $BelPl_{ij}(A,B) = Bel_i(A)\, Pl_j(B)$.

Definition 71 (The Judge)**.** A judge, or decision maker, or jury has to reach a verdict about the culpability or innocence of the suspects based on the available evidence and his own judgement. Mathematically, the judge is associated with a family of conditional weights which is a family of non-negative functions on $\mathcal{P}(X)$ conditioned by the available evidence.

For the body of evidence $\#i$ for which m_i is the probability (credibility) distribution induced on $\mathcal{P}(X)$ the judge must determine weights $w_i(\cdot \mid \cdot)$[1] where

$$w_i(\cdot \mid \cdot) : \mathcal{P}(X) \times F_i \longrightarrow [0, \infty) .$$

The judge assignes weight $w_i(C \mid A)$ which represents the culpability of the subset $C \in \mathcal{P}(X)$ if the i-th body of evidence focuses on the culpability of the subset $A \in F_i$. Here both C and A are a collections of suspects; A must be a focal set of body of evidence $\#i$ but C is arbitrary. The larger the weight the larger the potential culpability of the corresponding subset of suspects. From mathematical point of view, except nonnegativity, the only condition imposed on the family of weights is

$$\sum_{C \in P(X)} \sum_{A \in F_i} w_i(C \mid A) \; m_i(A) = 1 . \tag{13.7}$$

The judge may assign positive weights to many $w_i(C \mid \cdot)$, since, in his view, the credibility in many focal sets of of the body of evidence $\#i$ may transfer to the same $C \in \mathcal{P}(X)$. When we multiply the weight $w_i(C \mid \cdot)$ by $m_i(\cdot)$, and sum over focal sets of body of evidence $\#i$ we get a new credibiliy distribution on $\mathcal{P}(X)$ given by

$$\mu_i(C) = \sum_{A \in F(X; m_i)} w_i(C \mid A) \, m_i(A) , \tag{13.8}$$

abbreviated by $\mu_i = w_i \star m_i$. The larger the weight the larger the potential culpability of the corresponding subset of suspects. From mathematical point of view, except nonnegativity, the only condition imposed on the family of weights is

$$\sum_{C \in P(X)} \sum_{A \in F(X; m_i)} w_i(C \mid A) \, m_i(A) = 1, \tag{13.9}$$

The conditions on the weights of positivity and satisfying Eq. 13.9 guarantee that that μ_i given by Eq. 13.8 is a probability (credibility) distribution on $\mathcal{P}(X)$.

Definition 72. A family of weights is $probabilistic$ if they satisfy the equalities

$$\sum_{C \in P(X)} w_i(C \mid A) = 1 \text{ for every } A \in F(\; X; m_i) \; . \tag{13.10}$$

Obviously, (13.10) implies (13.9) but the converse is not necessarily true. If the family of weights is probabilistic and objective, based exclusively on relative frequencies, then $w_i(C \mid A)$ may be calculated using the standard formula for conditional probabilities. If, however, the family of weights is both nonprobabilistic and subjective,

[1] In probability theory, it is common to use a dot as a placeholder for an unnamed and unnumbered variable.

then $w_i(C \mid A)$ simply reflects what the judge believes about the culpability of C if the direct evidence focuses on the subset A and no special rules is necessarily used for getting it.

Some important kinds of judge weighting are:

Reliance If the judge fully relies on the i-th body of evidence, then $w_i(A \mid A) = 1$ and $w_i(C \mid A) = 0$ if C is different from A, for every $A \in F(X; m_i)$, which implies $\mu_i(A) = m_i(A)$.

Indifference If the judge focuses on $B \in P(X)$ regardless of what the i-th body of evidence says, then $w_i(B \mid A) = 1$ for every $A \in F(X; m_i)$, which implies $\mu_i(B) = 1$.

13.2.1. Generalization

The formulas above can be generalize easily to weighting joint credibility distributions. They can also be generalized to cases where the universes of the experts do not match. For example, Ann and Bob may arrive at different lists of subsets, Ann may focus on subsets of $X = \{Ralph, Sue, Tom\}$ while Bob focuses on subsets of $Y = \{Ralph, Sue, Ursula\}$. In addition, it may be that the Judge has his own list of suspects Z. We assign weights $w_{ij}(\,\mid\,,\,)$ to a mixed evidence inducting the joint credibility distribution m_{ij} on $\mathcal{P}(X) \times \mathcal{P}(Y)$. Thus the credibility distribution induced on $\mathcal{P}(X)$ by the weighted mixed (i,j)-the body of evidence is

$$\mu_{ij}(C) = \sum_{(A,B)\in F_{ij}} w_{ij}(C \mid A, B)\, m_{ij}(A, B) \text{ with } C \in \mathcal{P}(Z), \tag{13.11}$$

where $w_{i,j}(C \mid A, B)$ is the judge's weight of the subset $C \in \mathcal{P}(Z)$ given the mixed evidence $(A, B) \in F_{ij}$.

13.2.2. Fuzzy Evidence

Let $F : X \longrightarrow [0,1]$ be a fuzzy set. Then, the number $F(x)$ is the degree of membership of the element $x \in X$ to the fuzzy set F. While it is not necessary that $\{F(x) \mid x \in X\}$ is a probability distribution on X but, as shown in Guiasu (1993a) that it induces a probability distribution m_F on $\mathcal{P}(X)$, defined by

$$m_F(A) = \prod_{x\in A} F(x) \prod_{y\in A} [1 - F(y)] \text{ with } A \in P(X)\ . \tag{13.12}$$

We note that if $0 < X_F(x), 1$, for every $x \in X$, then $F = \mathcal{P}(X)$. All the consideration made above could be applied to the case when the available evidence is provided by fuzzy sets defined on X.

13.3. The Tuesday Night Club.

Here we present a long example first presented by Silviu Guiasu Guiasu (1993b, 1994), it is taken from Agatha Cristie's story "The Tuesday Night Club."

Breifly, Mrs Jones has died and the characters involved are:

J Mr. Jones, the husband of Mrs. Jones,

C Miss Clark, Mrs. Jones companion,

Dr The Doctor,

D The daughter of the Doctor,

G Gladys Lynch, the Maid.

Thus the universe is $U = \{J, C, Dr, D, G\}$.

There will be several bodies of evidence, each numbered $\#i$ and giving a credibility distribution $m_i : U \to [0, 1]$. Notationally, $m_i(\{C, D\})$ would be the credebility implied $(\Rightarrow)$ by the ith body of evidence to the proposition that the death of Mrs. Jones was perpetrated by Miss Clark and the Doctor's daughter. In addition $m_i(\emptyset)$ is the evidence that no one is guilty, and that the death was just an unfortunate acccident. As is usual in evidence theory, only the positive weights are mentioned in each body of evidence. All other values are zero.

Body of Evidence #1: J, C, and Mrs Jones sat down to a supper consisting of tinned lobster and salad, trifle, and bread and cheese. Later in the night all three became violenetly ill and the Doctor was summoned. Both J and C recovered, but Mrs. Jones died and was duly barried. The death certificate listed ptomaine poisoning as the cause of death.

$$\#1 \Rightarrow m_1(\emptyset) = 1$$

Body of Evidence #2: J had spent the previous night at a small hotel in Birmingham. The next morning the chambermaid found on the blotting paper the folllowing phrases; "entirely dependent on my wife," "when she is dead I will," and "hundreds of thousands." Also J had been very attentive to D. He also benefitted by his wifes death to the amount £8000.

$$\#2 \Rightarrow m_2(\{J\}) = 1$$

Body of Evidence #3: An autopsy was ordered. The body was exhumed and it was determined that Mrs. Jones had died of arsenic poisoning.

$$\begin{aligned}
\#3 \quad &\Rightarrow \quad m_3(\{J\}) = \frac{1}{5} \\
&\Rightarrow \quad m_3(\{C\}) = \frac{1}{5} \\
&\Rightarrow \quad m_3(\{Dr\}) = \frac{1}{5} \\
&\Rightarrow \quad m_3(\{D\}) = \frac{1}{5} \\
&\Rightarrow \quad m_3(\{G\}) = \frac{1}{5}
\end{aligned}$$

Body of Evidence #4: J's testimony. The freindship with D had ended two months before the death. The phrases on the blotter came from an innocent letter he had written to his brother.

$$\#4 \Rightarrow m_4(\emptyset) = 1$$

Body of Evidence #5: Dr's testimony plus an investigation performed upon him by Scotland Yard

$$\#5 \Rightarrow m_5(\emptyset) = 1$$

Body of Evidence #6: After supper, J had gone downstairs and demanded a bowl of corn-flour for his wife. This had been prepared by G. J had waited and carried the bowl up to his wife personally. J had motive and opportunity.

$$\#6 \Rightarrow m_6(\{J\}) = 1$$

Body of Evidence #7: C's testimony. The bowl of corn-flour was drunk by her. She was banting at the time and she was always hungry. and Mrs. Jones had changed her mind about tasting the corn-flour.

$$\#7 \Rightarrow m_7(\emptyset) = 1$$

Subsequently, Sir Henry of Scotland yard asked four consultants for their conclusions; Joyce, Mr. Pethric, Ramod, and Miss Marple. In the following computations we are using formula () and hence

$$m_{1,2,3,4,5,6,7} = m_1 m_2 m_3 m_4 m_5 m_6 m_7 \,.$$

Mr. Pethric (a solicitor, relying on facts and money): J was guilty and C sheltered him for money. She lied about the corn-flour to protect him. This gives five statements of the form

$$\Rightarrow w\left(\{J, C\} \mid \emptyset, \{J\}, A_3, \emptyset, \emptyset, \{J\}, \emptyset\right) = 1$$

where A_3 will cycle through the focal sets of the body of evidence introduced in $\#3$. Thus we replace A_3 with $\{J\}, \{C\}, \{Dr\}, \{D\}, \{G\}$ sequentially. Using these weights and the seven bodies of evidence in formula gives $\mu(J, C) = 1$.

Joyce (a young artist, relying on intuition): C was guilty. She was in love with J and hated his wife. As in the previous analysis we get five weights as A_3 is replaced by each of $\{J\}, \{C\}, \{Dr\}, \{D\}, \{G\}$.

$$\Rightarrow w\left(\{C\} \mid \emptyset, \{J\}, A_3, \emptyset, \emptyset, \{J\}, \emptyset\right) = 1$$

In this case we get the verdict implied is $\Rightarrow \mu(C) = 1$.

Raymond (a young writer, relying on imagination): D was guilty. After diagnosing the poisoning symptoms he sent a messenger home for some opium pills for Mrs. Jones, to relieve her acute pain. D, who was in love with J, had motive and opportunity. D therefore sent back pills containing arsenic.

$$\Rightarrow w\left(\{D\} \mid \emptyset, \{J\}, A_3, \emptyset, \emptyset, \{J\}, \emptyset\right) = 1$$

In this case we get the verdict implied is $\Rightarrow \mu(D) = 1$.

Miss Marple (an old lady relying on life experience and analogy): A similar case had happened in Mary St. Mead village. G performed the murder, goded by J who made her his murder instrument.

$$\Rightarrow w\left(\{G,J\} \mid \emptyset,\{J\},A_3,\emptyset,\emptyset,\{J\},\emptyset\right)=1$$

In this case we get the verdict implied is $\Rightarrow \mu(\{G,J\})=1$.

Scotland yard (gives each body of evidence equal weight): Scotland yard then evaluates each individual suspect, as well as the possibility that it was an accident. Now only the body of evidence $\#3$ has more than one focal set. When that focal set is $\{J\}$, that is when $A_3=\{J\}$, we are looking at something of the form $w\left(X \mid \emptyset,\{J\},\{J\},\emptyset,\emptyset,\{J\},\emptyset\right)$ which is only positive if $X=\emptyset$ or $X=\{J\}$. Notice that the weight is positive only if the set on the left of the "given" bar "|" is contained in the list on the right of the bar. In this particular case four of the bodies of evidence (numbers $\{\#1,\#4,\#5,\#7\}$) focus on the empty set and three (numbers $\{\#2,\#3,\#6\}$) focus on the set containing J. Therefore

$$w\left(\emptyset \mid \emptyset,\{J\},\{J\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{4}{7}$$

$$w\left(\{J\} \mid \emptyset,\{J\},\{J\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{3}{7}$$

When $A_3=\{C\}$ there are three different focal sets given, and we get the following weights:

$$w\left(\emptyset \mid \emptyset,\{J\},\{C\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{4}{7}$$

$$w\left(\{J\} \mid \emptyset,\{J\},\{C\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{2}{7}$$

$$w\left(\{C\} \mid \emptyset,\{J\},\{C\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{1}{7}$$

When A_3 cylcles through the remaining three focal sets, $\{Dr\}$, $\{D\}$, and $\{G\}$ we get results similar to the previos case. When $A_3=\{Dr\}$ we have:

$$w\left(\emptyset \mid \emptyset,\{J\},\{Dr\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{4}{7}$$

$$w\left(\{J\} \mid \emptyset,\{Dr\},\{C\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{2}{7}$$

$$w\left(\{Dr\} \mid \emptyset,\{J\},\{Dr\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{1}{7}$$

When $A_3=\{D\}$ we have:

$$w\left(\emptyset \mid \emptyset,\{J\},\{D\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{4}{7}$$

$$w\left(\{J\} \mid \emptyset,\{J\},\{D\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{2}{7}$$

$$w\left(\{D\} \mid \emptyset,\{J\},\{D\},\emptyset,\emptyset,\{J\},\emptyset\right)=\frac{1}{7}$$

Finally, when $A_3=\{G\}$ we have:

$$w(\emptyset \mid \emptyset, \{J\}, \{G\}, \emptyset, \emptyset, \{J\}, \emptyset) = \frac{4}{7}$$
$$w(\{J\} \mid \emptyset, \{J\}, \{G\}, \emptyset, \emptyset, \{J\}, \emptyset) = \frac{2}{7}$$
$$w(\{C\} \mid \emptyset, \{J\}, \{G\}, \emptyset, \emptyset, \{J\}, \emptyset) = \frac{1}{7}$$

The amalgamation of the conditional weights and the bodies of evidence, using formula (4) we produce the following results:

$$\begin{array}{c} \mu(\emptyset) = \frac{4}{7} \\ \mu(\{J\}) = \frac{11}{35} \\ \mu(\{C\}) = \frac{1}{35} \\ \mu(\{Dr\}) = \frac{1}{35} \\ \mu(\{D\}) = \frac{1}{35} \\ \mu(\{G\}) = \frac{1}{35} \end{array}$$

Body of Evidence #8: G's deathbed testimony. J had promised to marry her when his wife was dead. Following J's instruction she put arsenic on the trifle (thousands and thousands is UK slang for sugar sprinkles). Only Mrs. Jones ate the trifle since G was on a diet (banting) and J just brushed off the sprinkles. She had a child with J which died at birth and J deserted her for another woman.

Scotland yard (based on this confession): Now calculates

$$\Rightarrow w(\{G, J\} \mid \{G, J\}) = 1$$

and

$$\Rightarrow \mu(\{G, J\}) = 1$$

13.4. Special Cases

13.4.1. Hooper's Rule

Suppose that Ann and Bob both have bodies of evidence that divide the universe into A and $\bar{A} = X - A$. Thus $F_1 = F_2 = \{A, \bar{A}\}$ and the respective cridibilty distributions are m_1 and m_2 respectively. The judge decides that evidence fuses onto A whenever either Ann or Bob focus on A and that evidence fuses onto $\bar{A}$ only when both Ann and Bob agree on $\bar{A}$. The judge's weights are then

$$\begin{aligned} w_{1,2}(A \mid A, A) &= 1 \\ w_{1,2}(A \mid A, \overline{A}) &= 1 \\ w_{1,2}(A \mid \overline{A}, A) &= 1 \\ w_{1,2}(\overline{A} \mid \overline{A}, \overline{A}) &= 1\,. \end{aligned}$$

Then, according to Eq. 13.8, we have

$$
\begin{aligned}
m_{1,2}(A) &= 1 - [1 - m_1(A)]\,[1 - m_2(A)] \\
m_{1,2}(\overline{A}) &= m_1(\overline{A})m_2(\overline{A})\,.
\end{aligned}
$$

According to Lindley Lindley (1987), this rule for combining evidence was used by G. Hooper in 1685.

13.4.2. Dempster's Rule

Suppose that Ann and Bob have arbitrary bodies of evidence.

The judge adds up all the weights of evidence where Ann and Bob have some agreement, i.e., where A is a focal element for Ann and B is a focal element for Bob and $A \cap B \neq \emptyset$. The judge uses this weight as a normalizing factor so that

$$
w_{1,2}\left(A \cap B \mid A, B\right) = \left[1 - \sum_{C \in F_1, D \in F_2}^{C \cap D = \emptyset} m_1\left(C\right) m_2\left(D\right)\right]^{-1},
$$

for all $A \in F_1$, $B \in F_2$, $A \cap B \neq \emptyset$. With these weights Eq. 13.8, becomes

$$
\mu_{1,2}\left(C\right) = \frac{\sum\limits_{C = A \cap B} m_1\left(A\right) m_2\left(B\right)}{1 - \sum\limits_{A \cap B = \emptyset} m_1\left(A\right) m_2\left(B\right)}\,,
$$

which is Dempster's rule Dempster (1967b) of combining two independent bodies of evidence. It gives equal credit to the common evidence and discards any other evidence. According to Shafer (1976), in the special case of a universe containing only two elements, this rule was used by J.H. Lambert in his Neues Organon published in 1764.

13.4.3. Jeffrey's Rule

13.4.3.1. First interpretation

Suppose that Ann's information is a probability distribution p_Y on Y and that Bob's evidence is a probability distribution q_Z on Z. The judge will fuse evidence onto $X = Y \times Z$. We will use the shorthand notation of $y \times Z$ for the set of ordered pairs $\{y\} \times Z = \{\langle y, z_1\rangle, \langle y, z_2\rangle, \cdots, \langle y, z_n\rangle\}$, and, similarly, $Y \times z$ is shorthand for the set of pairs $Y \times \{z\} = \{\langle y_1, z\rangle, \langle y_2, z\rangle, \cdots, \langle y_m, z\rangle\}$ Assume that Ann's evidence is of the form

$$
\begin{gathered}
F_1 = \{y \times Z \mid y \in Y\}\,, \\
m_1\left(y \times Z\right) = p_Y\left(y\right),
\end{gathered}
$$

and Bob's evidence is of the form

$$
\begin{gathered}
F_2 = \{Y \times z \mid z \in Z\}\,, \\
m_2\left(Y \times z\right) = q_Z\left(z\right).
\end{gathered}
$$

Finally we will in addition assume that there exists a conditional probaility distribution $p_Z(\cdot \mid y)$ on Z given $y \in Y$, and p_Z, the *prediction* probability distribution on Z is defined by

$$p_Z(z) = \sum_{y \in Y} p_Z(z \mid y)\, p_Y(y),$$

where p_Y is interpreted as being the *prior* probability distribution on Y.

The judge fuses the evidence with weights

$$w(\{\langle y, z\rangle\} \mid y \times Z,\, z \times Y) = p_Z(z \mid y) / p_Z(z)\ . \tag{13.13}$$

giving

$$\mu_{1,2}(C) = \sum_{A \in F_1} \sum_{B \in F_2} w(C \mid A, B)\ m_1(A)\, m_2(B)\ .$$

Then, when C contains a single ordered pair, we conclude

$$\mu_{1,2}(\{\langle y, z\rangle\}) = \frac{p_Y(y)\, p_Z(z \mid y)}{p_Z(z)} q_z(z),$$

which is a probability distribution on $Y \times Z$. Its marginal probability distribution, namely

$$\begin{aligned} p_Y(y \mid q_Z) &= Bel_{12}(y \times Z) \\ &= \sum_{z \in Z} \mu_{1,2}(\{\langle y, z\rangle\}) \\ &= \sum_{z \in Z} \frac{p_Y(y)\, p_Z(z \mid y)}{p_Z(z)} q_z(z), \end{aligned} \tag{13.14}$$

which is Jeffrey's rule Jeffrey (1983) for calculating the *posterior* probability distribution on Y given the *actual* probability distribution q_Z on Z.

13.4.3.2. Second interpretation

Jeffrey's rule may be obtained more directly from Eq. 13.8, as a weighting with indirect evidence. Indeed, let X and Y be two finite crisp (Cantor) sets and m a probability distribution on $P(Y)$ such that

$$\begin{aligned} F &= \{\{y\} \mid y \in Y\}, \\ m(\{y\}) &= q(y), \end{aligned}$$

where q is the *actual* probability distribution on Y. Taking the only positive weights on $P(X)$ to be

$$w(\{x\} \mid \{y\}) = \frac{p(y \mid x)\, p(x)}{\sum_{x \in X} p(y \mid x)\, p(x)}$$

where p is a *prior* probability distribution on X and $p(\cdot \mid x)$ is a conditional probability distribution on Y given $x \in X$, the weighted probability distribution becomes

$$\begin{aligned} p(y \mid q_Z) &= \mu(\{x\}) && (13.15) \\ &= \sum_{y \in Y} w(\{x\} \mid \{y\})\, q(y) && (13.16) \\ &= \sum_{y \in Y} \frac{p(y \mid x)\, p(x)}{\sum_{x \in X} p(y \mid x)\, p(x)} q(y)\,, \end{aligned}$$

which is Jeffrey's rule for calculating the *posterior* probability distribution on X.

13.4.4. Bayes' Rule

Taking

$$q_Z(z) = \begin{cases} 1 & \text{if } z = z_0 \\ 0 & \text{if } z \neq z_0 \end{cases}$$

the formula 13.14becomes Bayes' rule for calculating the *posterior* probability distribution, namely

$$\begin{aligned} p_Y(y \mid z_0) &= p(y \mid q_Z) \\ &= \frac{p_Z(z_0 \mid y)\, p_Y(y)}{p_Z(z_0)}\,. \end{aligned}$$

Bayes' rule may also be obtained from Eq. 13.16 by taking q to be a degenerate probability distribution focussed on a single element $\{y_0\}$, i.e. $q(\{y_0\}) = 1$.

13.5. Conclusions

In Shafer's approach to evidence $m(\emptyset)$ has to be always equal to zero. This is an unnecessary restriction because m is not obtained by extending a probability distribution on X to a probability measure on $P(X)$, but is directly defined as a probability distribution on $P(X)$, in which case $m(\emptyset)$ could be positive, corresponding to the frequent case when there is a positive probability of having nobody guilty in the universe X.

"The process of reaching a verdict essentially depends on how the available evidence is used by the judge or jury. The evidence may be significant, partially relevant, or misleading and the judge may use it in an objective or subjective way. "

Homework

Let $X = \{x_1, x_2, x_3\}$ and let

Body of Evidence 1		
$m(\{x_1\})$	$=$	0.1
$m(\{x_3\})$	$=$	0.3
$m(\{x_1, x_2\})$	$=$	0.4
$m(X)$	$=$	0.2

Body of Evidence 2		
$m(\{x_1, x_3\})$	$=$	0.2
$m(\{x_2, x_3\})$	$=$	0.3
$m(\{x_1, x_2\})$	$=$	0.4
$m(X)$	$=$	0.1

Body of Evidence 3		
$m(\{x_1\})$	$=$	0.1
$m(\{x_1, x_2\})$	$=$	0.3
$m(X)$	$=$	0.6

Body of Evidence 4		
$m(\{x_3\})$	$=$	0.1
$m(\{x_2, x_3\})$	$=$	0.4
$m(X)$	$=$	0.2

Body of Evidence 5		
$m(\{x_1\})$	$=$	0.1
$m(\{x_2\})$	$=$	0.3
$m(\{x_3\})$	$=$	0.6

Body of Evidence 6		
$m(\{x_1\})$	$=$	0.7
$m(\{x_2\})$	$=$	0.3

1. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Dempster's Rule.
2. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Hooper's Rule.
3. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Jeffrey's Rule.
4. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Bayes's Rule.
5. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Dempster's Rule.
6. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Hooper's Rule.
7. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Jeffrey's Rule.
8. Fuse the data in *Body of Evidence 1* and *Body of Evidence 2* using Bayes's Rule.
9. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Dempster's Rule.
10. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Hooper's Rule.
11. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Jeffrey's Rule.
12. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Bayes's Rule.
13. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Dempster's Rule.
14. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Hooper's Rule.
15. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Jeffrey's Rule.
16. Fuse the data in *Body of Evidence 3* and *Body of Evidence 2* using Bayes's Rule.

14. Dempster $\neq$ Shafer

14.1. Introduction

In science, it is often usefull, and interesting, to read the papers where an idea originates.

Shafer's book on evidence theory recast Dempster's work into an elegant form. It providede many interesting additions and expansions. But did it capture all of the characteristics of Dempster's work?

Dempster's original formulation of upper and lower probabilities Dempster (1967b,a) was originally couched in terms of probability theory and inverse mappings. Dempster considered the situation where there was a finite set $X$with an unknown probability distribution p. What was known was the probability measure of a certain class of events, that is, there were sets $A_1, A_2, \ldots \subseteq X$ where $P(A_i)$ was known. Given this data Dempster asked "What can be inferred about the probability of an arbitrary event?"

Dempster stated that the data imparted constraints upon the unknown probability distribution. One potential outcome was that the data was inconsistent with any probability distribution. The second potential was that only one probability distribution fit the known data. The third potential outcome was that their existed a sheaf of probability distributions that fit the data.

Dempster examined the third potential outcome, and determined that for an arbitrary event, the best we could do was determine an upper rand lower probability, and that the true probability was bounded by these two vales.

Shafer wrote his treatise on Evidence theory using Dempster's work as an inexact basis. Instead of staring from a known set of data about certain events, Shafer starts with a probability distribution m upon the power set of the (finite) universe of discourse. He terms this distribution a basic probability assignment, or **bpa**. Using the **bpa** he shows how to determine an upper and lower measure upon any subset of the universe and, calls these measures Plausibility and Belief.

14.1.1. Upper and Lower Probabilities

In probability theory, weights expressing evidential claims are assigned to individual elements of some universal set X. The probability of any subset of the sample space X is calculated by adding the weights of the subset's elements. Dempster examined the inverse problem. Suppose you know the probabilities of some subsets of a universal set. What can you say about the probabilities of the other subsets and of the singletons?

If the evidence is not contradictory then, usually, the problem does not allow for an exact solution. Instead, for each subset of the universe, a maximum and minimum probability consistent with the given probabilities can be calculated. The correct solution then lies somewhere in the interval between these two values.

The minimum consistent probability assigned to a set $A \subseteq X$ is called the lower probability of A and is denoted $P_*(A)$. The maximum consistent probability assigned to a set $A \subseteq X$ is called the upper probability of A and is denoted $P^*(A)$. The correct probability assigned to A, $Pro(A)$ must be bracketed by these values, i.e.,

$$P_*(A) \le Pro(A) \le P^*(A) \tag{14.1}$$

or

$$Pro(A) \in [P_*(A), P^*(A)] \quad . \tag{14.2}$$

As expressed by Dempster, there is a family of probability measures, $\mathbb{P}$, bounded by the dual upper and lower probabilities and consistent with the given evidence,

$$\mathbb{P} = \{Pro \,|\, P_*(A) \le Pro(A) \le P^*(A) \text{ for all } A \subseteq X\} \quad . \tag{14.3}$$

14.2. Shafer not equal to Dempster

While Shafer's Evidence theory has a beautiful formulation, as we will see, it does not capture all of the possibilities inherent in Dempster's original formulation.

In Dempster's framework we have a universe and a collection of observed probabilities for events in this universe. If the data is non-condradictory there are usually many probability distributions on the universe that could produce the given observed data. For every event a largest and smallest probability consistent with the data can be calculated. Dempster termed these values the lower probabilities, *LP*, and upper probabilities, *UP*.

Example 91. Let $X = \{a, b, c, d\}$ with given data $P(a, b) = P(b, c) = \frac{2}{3}$. It is not to difficult to derive that the distributions on X that could reproduce the data are of the form $\mathbb{P} = \left\langle \frac{2}{3} - \alpha, \alpha, \frac{2}{3} - \alpha, \alpha - \frac{1}{3} \right\rangle$ where $\alpha \in \left[\frac{1}{3}, \frac{2}{3}\right]$. Given these distributions, maximal and minimal probabilities for all events can be calculated. They are given in the following table. In addition, using the formula $m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathrm{Bel}(B)$, and associating the lower probability of Dempster with the Belief measure of Shafer, $Bel = P_*$, we calculate the basic probability assignment in the last column of Table 14.1.

This gives $m(b, c, d) = \frac{-1}{3}$, a great difficulty. The **bpa** does still have a sum equal to one. However we haven negative weights associated with certain sets. This contradicts a basic axiom of evidence theory as well as its standard interpretation as a probability distribution upon the power set of X, excluding the empty set. We see then that Shafer's interpretation is not an all-inclusive one.

14.3. Negative evidence and quantum mechanics

Quantum mechanics is an extremely successful theory in physics for the prediction of the microverse. The Wigner function which corresponds to a probability distribution is unfortunately negative in certain regions. Evidence theory, as originated by Dempster, corresponds in some cases to a negative basic probability assignment, a result

a	b	c	d	P	P_*	P^*	$m?$
0	0	0	0	0	0	0	.
0	0	0	1	$\alpha - \frac{1}{3}$	0	$\frac{1}{3}$	0
0	0	1	0	$\frac{2}{3} - \alpha$	0	$\frac{1}{3}$	0
0	0	1	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
0	1	0	0	α	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
0	1	0	1	$2\alpha - \frac{1}{3}$	$\frac{1}{3}$	1	0
0	1	1	0	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
0	1	1	1	$\frac{1}{3} + \alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	0	0	0	$\frac{2}{3} - \alpha$	0	$\frac{1}{3}$	0
1	0	0	1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
1	0	1	0	$\frac{4}{3} - 2\alpha$	0	$\frac{2}{3}$	0
1	0	1	1	$1 - \alpha$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{-1}{3}$
1	1	0	0	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$
1	1	0	1	$\frac{1}{3} + \alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	1	1	0	$\frac{4}{3} - \alpha$	$\frac{2}{3}$	1	$\frac{-1}{3}$
1	1	1	1	1	1	1	$\frac{2}{3}$

Table 14.1.: The Basic Probability Assignment, Lower and Upper probabilities of Example 91.

that Shafer disallows. This paper shows that allowing negative evidence permits the modeling of interference effects in the two slit experiment.

14.4. Quantum mechanics and negative probability

Quantum mechanics is one of the most successful physical theories in the history of science. Quantum mechanics predictions about the results of an experiment are unerringly accurate. However, quantum mechanical predictions about the result of an experiment are probabilistic in nature. Quantum mechanics cannot predict whether or not a particle will decay at a specific moment. Quantum mechanics can predict what percentage of a mass of particles will decay over a given time period with spectacular accuracy.

One of the many interesting results in quantum mechanics are negative probabilities Feynman (1987). Feynman states that the probability of an event that can actually occur never turns out to be negative, but that the probabilities in the intermediate calculations can have negative values.

For example suppose an apples vendor starts with 6 apples. At noon he is resupplied with 5 additional apples. During the day he sells a total of 8 apples. The result is $6 - 8 + 5$ apples, or 3 apples. An intermediate calculation if we group the terms appropriately gives $6 - 8 = -2$ apples or negative two apples. Now, no one can have a negative number of apples since this is simply the result of a bookkeeper, starting with an initial stock, first subtracting the sales, and then adding the restock. The bookkeeper's intermediate results are negative but this is not reflected in the real world. Another example, from statistical signal processing, occurs when we take a

A	H	T
H	α	$\frac{1}{3}-\alpha$
T	$\frac{1}{3}-\alpha$	$\frac{1}{3}+\alpha$

B	H	T
H	$\frac{1}{9}$	$\frac{2}{9}$
T	$\frac{2}{9}$	$\frac{4}{9}$

C	H	T
H	0	p
T	$1-p$	0

D	H	T
H	$\frac{1}{6}$	$\frac{1}{2}$
T	$\frac{1}{2}$	$\frac{-1}{6}$

Table 14.2.: Joint probabilities of the outcomesd of tossing two coins. (A) Joint probailities induced by the marginal ditributions of $P(H) = \frac{1}{3}$ and $P(T) = \frac{2}{3}$ where α ranges from 0 to $\frac{1}{3}$. (B) The noninteractive solution of (A). (C) Marginal probabilites where the $P(H) = P(T) = 0$ (C) Pseudo probabilites that satisfy both the requirements of both (A) and (C).

function g and use its Fourier transform G (which contains $\sin$ and $\cos$ terms which may be negative in regions) in processing to remove noise. In the end we inverse-transform the composition of G and the filter F to recover the source minus the noise.

A very popular and accessible review of Feynman's paper was published in Dietrich Leibfried and Monroe (1998). It gives a simplified example of the troubles probability theory has when applied to quantum mechanical applications. Imagine we have two coins. If either coin is flipped independently in a black box, then the ratio of heads to tails is 1:2. For either coin we have that the probability of heads is one-third and the probability of tails is two-thirds, or $P(H_i) = \frac{1}{3}$ and $P(T_i) = \frac{2}{3}$ for $i = 1, 2$. However when we flip both coins together in the black box it is always the case that the coin one is heads and coin two is tails or coin one is tails and coin two is heads. That is, the events that both coins are heads or both coins are tails never occurs. Thus the probabilities of the events HH, HT, TH, and TT, are such that $P(HH,TT) = 0$ and $P(HT,TH) = 1$. Table 14.2 shows the results of various thinking. The marginals induce the probability distribution shown in Table (14.2.A) where α is in the interval 0 to $\frac{1}{3}$. The Table (14.2.B) shows the noninteractive solution where $P(HH) = P(H_1) * P(H_2)$, etc. The Table (14.2.C) shows a probability distribution that satisfies the requirements that HH and TT never occur together. The only way to get a solution to both requirements is to abandon positivity, a result shown in Table (14.2.D).

Lowe Lowe (1998) mentions that negative probabilities also appear in neural networks. Using a probabilistic learning rule that conditions individual neurons to learn, it turns out that some inner nodes have negative weights but that the output nodes always end up producing positive results. He also mentions that it is impossible, see Rosenblatt (1956) and Yamato (1971), to generate approximations to (unknown) pdfs which simultaneously satisfy the three 'axioms':

- Realness
- Positivity
- Reproduce correct expectation values.

If Lowe's proposition is correct then the first (the values are real numbers) and the third (the results correspond to the experiment) properties are not disposable, and positivity is the property we must abandon.

14.5. Negative evidence

A revised theory of evidence that allows for a negative **bpa** does not allow for negative beliefs. Here we have a terminology problem, beliefs and plausibilities are Lower and Upper Probabilities respectively and cannot be negative. While termed a **bpa** by Shafer, the function m that can be calculated from Bel using the formula in Eq. (12.4) is simply the combinatorial result of the Mobius inversion formula. A better name for the function m then is a basic evidential weight, or **bew**, which in our case is a function $m : \mathcal{P}(X) \to [-1, 1]$,which is required to satisfy three conditions:

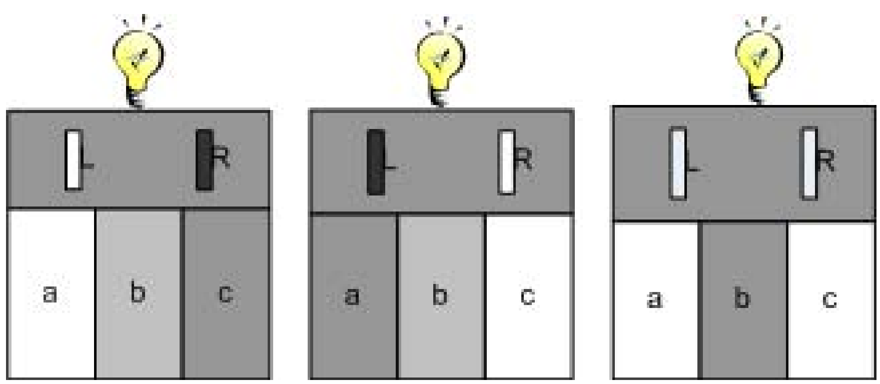
$$\begin{array}{rl} (i) & m(\emptyset) = 0 \\ (ii) & \sum_{A \in \mathcal{P}(X)} m(A) = 1 \\ (iii) & 0 \leq \sum_{B \subseteq A} m(B) \leq 1 \quad \forall A \subseteq X \end{array}$$

Under these new restrictions both Bel and Pl are still fuzzy measures, that is monotone, positive set functions.

It is not always easy to create **bew**s that satisfy requirement $\mathrm{(iii)}$. However, if we have start from a collection of lower probabilities, P_*, as in Dempster and Example (91), we can always calculate a **bew** using Eq. (12.4), with the assumption that $\mathrm{Bel} \equiv P_*$, that satisfies all the requirements of a revised evidence theory.

14.6. Quantum mechanics and negative evidence

If we abandon the positivity requirement for evidence theory and instead use a **bew** that ranges over the interval $[-1, 1]$ then we arrive at a theory that has some interesting applications. Consider for example the two slit experiment in physics where a polarized light is behind a screen containing two slits.



If only the left slit L is open then the photons spread out from it with a distribution that depends only on the distance from the L opening. Similarly if only the right slit R is open then the photons spread out from it with a distribution that depends only on the distance from the R opening. However if both slits are open then a curious interaction in the wave nature of particles (and the particle-wave model is at the heart of quantum mechanics) causes dark and light bands. If we consider a dark region,

b, in between the two slits then the probabilities of a particles landing there when either of the two slits is open is some positive probability β, or $P(b|L) = P(b|R) = \beta$. However the probability of a photon landing in b if both the slits are open is zero, so that $P(b|L,R) = 0$. This defies the monotone nature of probabilities. If we abandon positivity in evidence theory we can give a fascinating model of the interaction of the photons as presented in Table (14.3). The **bew** still sums to one, and note that all the lower probabilities or Beliefs are positive. The interesting **bew**s are the values $m(b|L) = m(b|R) = 0.1$ and $m(b|L,R) = -0.2$. We can interpret $m(b|L,R) = -0.2$ as saying that when slits L and R are both open then the weight that flows to event $(b|L,R)$ from $m(b|L)$ and $m(b|L)$ is canceled out. The corresponding Beliefs are $Bel(b|L) = Bel(b|R) = 0.1$ and $Bel(b|L,R) = 0.0$. All of the **bew**s and $\mathrm{Bel}s$ are spelled out in Table (14.3).

Set	**Slits**	**bew**	**Bel**
a	L	0.2	0.2
b	L	0.1	0.1
c	L	0.05	0.05
a,b	L	0	0.3
a,c	L	0	0.25
b,c	L	0	0.15
a,b,c	L	0	0.35
a	R	0.05	0.05
b	R	0.1	0.1
c	R	0.2	0.2
a,b	R	0	0.15
a,c	R	0	0.25
b,c	R	0	0.3
a,b,c	R	0	0.35
a	L,R	0.25	0.5
b	L,R	-0.2	0
c	L,R	0.25	0.5
a,b	L,R	0	0.5
a,c	L,R	0	1
b,c	L,R	0	0.5
a,b,c	L,R	0	1

Table 14.3.: Evidence ditribution that produces the desired lower Beliefs

Remark 9. It is not possible to get a better model of the coin flipping example using evidence theory. The marginal constraints combined with the double flipping constraints allow for a single probability distribution, the one given in Table (14.2.D). For this example the upper and lower probabilities are simply the probabilities $P \equiv P^* = P_*$ and m is identical to the single quasi-probaility distribution given in Table (14.2.D).

14.7. Conclusions

In Dempster's original papers, lower probabilities are the $\inf$ of the probability of an event over a constrained set of probability distributions. We have seen that a Mobius inversion of the lower probabilities can produce a function m that contains values that are negative. Negative values in this function can be interpreted as restrictions on the flow of probability weights onto an event.

Negative evidential weights may be useful in modeling other events where there are simultaneity restrictions. For example, the Heisenberg Uncertainty Principle states that it is impossible to simultaneously measure the mass and momentum of a particle to arbitrary degree of precision.

We should mention that there is much disagreement in evidence theory about the process of conditioning or information fusion, see for example Friedman and Halpern (1999).

Part IV.

Fuzzy Logic and Control

15. Fuzzy Logic

Classical logic is briefly described in Section 3. Implication, if A then B, is modeled in classical logic with the right arrow operator $A \rightarrow B$. The result of implication for every possible truth assignment is given in the Truth Table (15.1), where 0 represents false and 1 represents true.

Logic is one of those subjects that just invites argument in every sense of the word. Aristotle said "A truth can only be derived from previously known truths." The problem is, where do we start? How do we get first truths? In mathematics these known truths are called axioms. They are reasonable statements that can be accepted without justification. Except, every axiom of every formal system has been justified in numerous texts. And some axioms, such as the axiom of choice, that one can randomly choose an example element from a set, which is absolutely indispensable to most of mathematics, leads to conclusions that are not easily accepted (the well ordering theorem). "Where logic deals with ideals and abstractions it can have no meaning" This is a direct quote from one of the most famous logicians who ever lived, Bertram Russel. Thus classical logic is abstract and its results are meaningless — **literally meaningless.**

Numbers for example have no physical representation. There can be two apples but there is never a physical two. An apple is a thing you can see and feel and eat, a two cannot be seen, felt, tasted, etc. it can only be imagined. The variable v can be two but can v be an apple? No, a map is not a country, a picture is not a mountain, a poem is not a tree. There are even worse problems with logic, Godël *proved* that you could not prove all the truths that a formal logic could express, unless the logic contained paradox, a statement the was both true and false like "Everything I say is a lie".

The purpose of this chapter is to extend classical logic to fuzzy logic as classical or crisp sets were extended to fuzzy sets. However, as will be seen we may have to abandon the truth table () in its entirety. It is also important to note that this is not the first attempt to extend classical logic. Many others have tried, including Lukasiewicz who developed a three valued logic for true, false, and unknown as well as multi-valued

	A	B	$A \rightarrow B$
row 1	0	0	1
row 2	0	1	1
row 3	1	0	0
row 4	1	1	1

Table 15.1.: The truth table for the logical connective: implication.

logics. The logic of two-values is often called Boolean logic.

15.1. Implication

Fuzzy logic expands classical logic by simple expanding the role of the characteristic or membership function. Logical connectives are functions like $\max : [0,1]\times[0,1] \to [0,1]$ for or and $\min : [0,1] \times [0,1] \to [0,1]$ for and. But the point of fuzzy logic is not the abstract manipulation of truth values. It is not designed to serve in the construction of mathematical proofs, like "There are an infinite number of primes." The design of fuzzy logic is to build a better reasoning methodology for vague information.

If p is a proposition of the form "u is A" where A is a fuzzy set, for example, big pressure and q is a proposition of the form "v is B" for example, small volume then one encounters the following problem: How to define the membership function of the fuzzy implication $p \to q$ or, equivalently, $A(u) \to B(v)$? It is clear that $(A \to B)(u, v)$ should be defined pointwise i.e. $(A \to B)(u, v)$ should be a function of $A(u)$ and $B(v)$ since this is the approach for every other operator of fuzzy set theory, 6.

That is, there should exist a fuzzy implication operator $\mathsf{i} : X \times Y \to [0, 1]$ that allows one to calculate the value of $(A \to B)(u, v)$ given the values of of $A(u)$ and $B(v)$,

$$(A \to B)(u, v) = \mathsf{i}(A(u), B(v)). \tag{15.1}$$

We shall also use the notation

$$(A \to B)(u, v) = A(u) \to B(v) \tag{15.2}$$

to represent the fuzzy set that corresponds to the implication if u is A then v is B, where A and B are fuzzy sets on X and Y respectively.

In our interpretation $A(u)$ is considered as the truth value of the proposition "u is big pressure", and $B(v)$ is considered as the truth value of the proposition "v is small volume".

$$u \text{ is big pressure} \to v \text{ is small volume} \equiv A(u) \to B(v) \tag{15.3}$$

It should be remarked that most of the implication operators owe their genesis to logical identities. Of course using classical logical identities may be the wrong way to approach the problem of designing an implication operator. This is because many results of classical logical are not true for fuzzy set theory. For example the law of the excluded middle states that a statement must be true or false. In fuzzy logic a statement can be 0.7 true and 0.3 false. Remember, logic is a tool constructed for a purpose, it woks well at that purpose, as digital technology demonstrates conclusively. However analog technology allows for an infinite amount of values, and the world is analog (light, sound, etc.)

One possible extension of material implication to implications with intermediate truth values is

$$A(u) \to B(v) = \begin{cases} 1 & \text{if } A(u) \le B(v) \\ 0 & \text{otherwise} \end{cases} \tag{15.4}$$

This implication operator is called *Standard Strict* .

Example 92. Suppose that we have fuzzy sets A (representing big pressure) and B

(representing small volume) with $A(1) = 0.5 \rightarrow$ and $B(2) = 0.8$ then

$$\begin{aligned} 1 \text{ is big pressure} \rightarrow 2 \text{ is small volume} &= A(3) \rightarrow B(1) \\ &= 0.5 \rightarrow 0.8 \\ &= 1 \end{aligned} \tag{15.5}$$

However,it is easy to see that this fuzzy implication operator is not appropriate for real-life applications. Namely,let $A(2) = 0.8$ and $B(2) = 0.8$. Then we have

$$A(2) \rightarrow B(2) = 0.8 \rightarrow 0.8 \tag{15.6}$$
$$= 1 \tag{15.7}$$

Let us suppose that there is a small error of measurement or small rounding error of digital computation in the value of $B(2)$,and instead 0.8 we have to proceed with 0.7999.

Then from the definition of Standard Strict implication operator it follows that

$$A(2) \rightarrow B(2) = 0.8 \rightarrow 0.7999 \tag{15.8}$$
$$= 0 \tag{15.9}$$

This example shows that small changes in the input can cause a big deviation in the output, i.e., our system is very sensitive to rounding errors of digital computation and small errors of measurement. In general, fuzzy systems are designed to be insensitive to small changes in membership degree. This means that fuzzy systems are designed to be robust.

A smoother extension of material implication operator can be derived from a trivial consequence of the isomorphism between classical logic and classical set theory. When A, B and C are classical sets in a universe X, then the following equivalence can be demonstrated

$$A \rightarrow B \equiv \sup \{C \mid A \cap B \subset C\} \ . \tag{15.10}$$

where $\sup$ is by set inclusion.

Using the above equivalence one can define the following fuzzy implication operator, where $w \in [0, 1]$

$$A(u) \rightarrow B(v) = \sup \{w \mid \min\{A(u), w\} \leq B(v)\} \tag{15.11}$$

that is,

$$A(u) \rightarrow B(v) = \begin{cases} 1 & \text{if } A(u) \leq B(v) \\ B(v) & \text{otherwise} \end{cases} \tag{15.12}$$

This operator is called *Gödel implication*. Using the definitions of negation and union of fuzzy subsets the material implication operator $p \rightarrow q = \neg p \vee q$ can be extended to fuzzy sets with the following definition

$$A(u) \rightarrow B(v) = \max\{1 - A(u), B(v)\} \tag{15.13}$$

This operator is called *Kleene-Dienes implication*.

In many practical applications one uses *Mamdani's* implication operator to model causal relationship between fuzzy variables. This operator simply takes the minimum of truth values of fuzzy predicates

$$A(u) \to B(v) = \min\{A(u), B(v)\} \tag{15.14}$$

It is easy to see this is not a correct extension of material implications, because $0 \to 0$ yields zero while the truth table for classical logic gives the truth value of $0 \to 0$ as 1. However, in the knowledge-based systems that will be demonstrated in the following chapters, rules for which the antecedent, the A part, are false are simply ignore. This is like a rule that says "If it is raining, then open your umbrella." If it is not raining we don't even think about this rule and could care less what it says to do. That is because fuzzy logic is for application not for symbolic manipulation. Symbolic manipulation is important, without it we would not have digital computers, however, symbolic manipulation is not the whole of thinking, though some have claimed that it is.

There are three important classes of fuzzy implication operators:

- s-implications: defined by

  $$a \to b = \mathsf{s}(\mathsf{c}(a), b) \tag{15.15}$$

  where s is a t-conorm (or s-norm, hence the name, s-implication) and c is a negation operator on $[0, 1]$. These implications arise from the Boolean formalism

  $$a \to b \equiv \neg a \vee b\ . \tag{15.16}$$

  Typical examples of s-implications are the Lukasiewicz and Kleene-Dienes implications.

- R-implications: obtained by residuation of continuous t-norm t, i.e.,

  $$a \to b = \sup\{c \in [0, 1] \mid \mathsf{t}(a, b) \leq c\} \tag{15.17}$$

  These implications arise from the Intutionistic Logic formalism. of Eq. (15.10) Typical examples of R–implications are the Gödel and Gaines implications.

- t-norm implications: if t is a t-norm then

  $$a \to b = \mathsf{t}(a, b)\ . \tag{15.18}$$

  t-norm implications are used as model of implication in many applications of fuzzy logic. These implication operators do not verify the properties of material implication, specifically, for any t-norm $\mathsf{t}(0, 0) = 0$ but in logic $0 \to 0 = 1$. Typical examples of t-norm implications are the Mamdani ($a \to b = \min\{a, b\}$) and Larsen ($x \to y = xy$) implications.

The most often used fuzzy implication operators are listed in the following table.

Name	**Label**	**Definition of** $a \to b$
Early Zadeh	i_m	$a \to b = \max\left[1-a, \min\left[a,b\right]\right]$
Lukasiewicz	i_a	$a \to b = \min\left[1, 1-a+b\right]$
Mamdani*	i_{mm}	$a \to b = \min\left[a,b\right]$
Larsen	i_l	$a \to b = ab$
Standard Strict	i_{ss}	$a \to b = \begin{cases} 1 & \text{if } a \le b \\ 0 & \text{otherwise} \end{cases}$
Gödel	i_g	$a \to b = \begin{cases} 1 & \text{if } a \le b \\ b & \text{otherwise} \end{cases}$
Gaines–Rescher	i_{gr}	$a \to b = \begin{cases} 1 & \text{if } a \le b \\ 0 & a > b \end{cases}$
Goguen	i_{gg}	$a \to b = \begin{cases} 1 & \text{if } a \le b \\ \frac{b}{a} & \text{otherwise} \end{cases}$
Kleene-Dienes	i_{kd}	$a \to b = \max\left[1-a, b\right]$
Reichenbach	i_r	$a \to b = 1{-}a{+}ab$
Yager	i_y	$a \to b = \begin{cases} 1 & \text{if } a = b = 0 \\ b^a & \text{otherwise} \end{cases}$

Table 15.2.: Fuzzy implication operators.

15.1.1. Examples

Example 93. Let $X = \{1, 2, 3\}$ and $Y = \{1, 2, 3\}$. Let us define two fuzzy sets, first A (big pressure) on X

$$A = \frac{0.5}{1} + \frac{0.8}{2} + \frac{1.0}{3}$$

and second B (small volume) on Y

$$B = \frac{1.0}{1} + \frac{0.8}{2} + \frac{0.2}{3} \ .$$

Early Zadeh implication uses the formula $a \to b = \max\{1-a, \min(a,b)\}$ so that the truth value for 1 is big pressure$\to$ 1 is small volume

$$\begin{aligned}\text{if } 1 \text{ is } A \text{ then } 1 \text{ is } B &= A(1) \to B(1) \\ &= 0.3 \to 1.0 \\ &= \max\left(1-0.3, \min\left[0.3, 1\right]\right) \\ &= 0.7 \ .\end{aligned}$$

The truth value for

$$\begin{aligned}A(1) \to B(2) &= 0.3 \to 0.8 \\ &= \max\left(1-0.3, \min\left[0.3, 0.8\right]\right) \\ &= 0.7\end{aligned}$$

and for

$$\begin{aligned} A(3) \to B(2) &= 1.0 \to 0.8 \\ &= \max(1 - 1.0, \min[1.0, 0.8]) \\ &= 0.8 \end{aligned}$$

All of these results can be summarized in the following table:

Early Zadeh

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.3	0.7	0.7	0.7
$A(2)$	0.8	0.8	0.8	0.2
$A(3)$	1.0	1.0	0.8	0.2

Here are the results of applying the formulas for the other implication operators to the fuzzy sets A and B:

Lukasiewicz

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	1.0	1.0	0.7
$A(2)$	0.8	1.0	1.0	0.4
$A(3)$	1.0	1.0	0.8	0.2

Mamdani

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	0.5	0.5	0.2
$A(2)$	0.8	0.8	0.8	0.2
$A(3)$	1.0	1.0	0.8	0.2

Larsen

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	0.5	0.5	0.2
$A(2)$	0.8	0.8	0.8	0.2
$A(3)$	1.0	1.0	0.8	0.2

Standard Strict

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	0.0	0.0	0.0
$A(2)$	0.8	0.0	1.0	0.0
$A(3)$	1.0	1.0	0.0	0.0

Godel

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	1.0	0.8	0.2
$A(2)$	0.8	1.0	1.0	0.2
$A(3)$	1.0	1.0	0.8	0.2

Goguen

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	1.0	1.0	0.4
$A(2)$	0.8	1.0	1.0	0.250
$A(3)$	1.0	1.0	0.8	0.2

Kleene-Dienes

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	1.0	0.8	0.5
$A(2)$	0.8	1.0	0.8	0.2
$A(3)$	1.0	1.0	0.8	0.2

Reichenbach

		$B(1)$	$B(2)$	$B(3)$
	$\to$	1.0	0.8	0.2
$A(1)$	0.5	1.0	0.9	0.6
$A(2)$	0.8	1.0	0.840	0.360
$A(3)$	1.0	1.0	0.8	0.2

Yager

		$B(1)$	$B(2)$	$B(3)$
	$\rightarrow$	1.0	0.8	0.2
$A(1)$	0.5	1.0	0.894	0.447
$A(2)$	0.8	1.0	0.837	0.276
$A(3)$	1.0	1.0	0.8	0.2

15.2. Axioms for Implication Operators

There have been attempts to characterize fuzzy implication operators axiomatically. These attempts are very similar in flavor to those used to characterize fuzzy and, or, and not as t–norms, t–conorms and complements. Here are the usual set of axioms for a fuzzy implication operator.

Axiom 1 (monotonicity in first argument). $a \leq b$ *implies that* $\mathsf{i}(a, x) \geq \mathsf{i}(b, x)$.

Axiom 2 (monotonicity in second argument). $a \leq b$ *implies that* $\mathsf{i}(x, a) \leq \mathsf{i}(x, b)$.

Axiom 3 (dominance of falsity). $\mathsf{i}(0, b) = 1$.

Axiom 4 (neutrality of truth). $\mathsf{i}(1, b) = b$.

Axiom 5 (identity). $\mathsf{i}(a, a) = 1$.

Axiom 6 (exchange). $\mathsf{i}(a, \mathsf{i}(b, x)) \leq \mathsf{i}(b, \mathsf{i}(a, x)$.

Axiom 7 (boundary condiition). $\mathsf{i}(a, b) = 1$ *if and only if* $a \leq b$.

Axiom 8 (contraposition). $\mathsf{i}(a, b) = \mathsf{i}(\mathsf{c}(b), \mathsf{c}(a))$ *for some fuzzy complement operator* c.

Axiom 9 (continuity). $\mathsf{i}(a, b)$ *is a continuous function of its arguments* a *and* b.

It is shown in Smets and Magrez (1987) that any implication operator that satisfies the above axiom schema can be characterized by a strictly increasing continuous function $f : [0, 1] \to [0, \infty)$ such that $f(0) = 0$. Since f is strictly increasing and continuous it has an inverse f^{-1} and the implication operator it characterizes is given by the formula

$$\mathsf{i}(a, b) = f^{-1}\left(f(1) - f(a) + f(b)\right)$$

and the complement operator that satisfies Axiom (8) is given by

$$\mathsf{c}(a) = f^{-1}\left(f(1) - f(a)\right) \ .$$

Example 94. $f(x) = x$. Then $\mathsf{i}(a, b) = \min\left[1, 1 - a + b\right]$ which is the Lukasiewicz implication operator and the complement operator is $\mathsf{c}(a) = 1 - a$.

The above example implies that the standard implication operator of fuzzy set theory should be the Łukasiewicz implication since it corresponds to the standard complement operator $c(a) = 1 - a$.

Example 95. $f(x) = x^w$ with $w \in [-1, \infty)$. Then $\mathsf{i}(a, b) = \min\left[1, \sqrt[w]{1 - a^w + b^w}\right]$ which is called a pseudo-Łukasiewicz implication operator and the complement operator is $\mathsf{c}(a) = \sqrt[w]{1 - a^w}$.

Example 96. $f(x) = \ln(1+a)$ with pseudo-inverse

$$f^{-1}(x) = \begin{cases} e^x - 1 & 0 \le x \le \ln 2 \\ 1 & otherwise \end{cases} .$$

Then

$$\mathsf{i}(a,b) = \min\left[1, \frac{1-a+2b}{1+a}\right]$$

which is called a Sugeno Type 1 implication operator and the complement operator is

$$\mathsf{c}(a) = \frac{1-a}{1+a} .$$

The axioms for implication operators presented above are not independent. For example the axioms of identity and dominance of falsity imply the axiom of boundary condition. However some of the proposed implication operators satisfy a restricted subset of all the axioms. Mamdani implication, for instance, has $\mathsf{i}(0,0) = 0$, which violates many of the axioms, yet this is probably the most applied of all the implication operators.

15.3. Approximate Reasoning

Remember (see Sec. (3)) that the main point of logic is deduction schemes.

The first problem with developing an applied fuzzy logic is to recognize that literal interpretation of symbolic logic is not going to work. Consider for example $(A \wedge (A \to B)) \to B$ which is the modus ponens rule written as a single logical statement. In two valued logic a statement that is always true is called a tautology and $(A \wedge (A \to B)) \to B$ is a tautology. But this expression does not truly capture a deduction scheme because it does not allow for the creation of free standing conclusions. The conclusion of the modus ponen rule of inference in tableau form is B a single new tautological statement.

Another problem with the tautology schemes is that we cannot directly adapt them to fuzzy set theory. For instance, let A be a fuzzy set on X, B be a fuzzy set on Y, and C be a fuzzy set on Z. If we examine the left hand side of Eq. (3.9) it contains the expression $((A \to B) \wedge (B \to C))$. Now by the constructions provided in Chapter (3) $A \to B$ and $B \to C$ are represented by fuzzy sets constructed from the linguistic variables (fuzzy sets) for A, B, and C using a selected implication operator. The fuzzy set (or fuzzy relation) for $A \to B$ is defined upon the space $X \times Y$ and the fuzzy set (or fuzzy relation) for $B \to C$ is defined upon the space $Y \times Z$. The standard interpretation of $\wedge$ is `and` which is processed using an intersection operator. But we cannot intersect the fuzzy sets $A \to B$ and $B \to C$ because they do not have the same domain! The same problem will occur if we try to process the left hand side of Eq. (3.5), $A \wedge (A \to B)$, which again attempts to form the intersection of two fuzzy sets on different domains. This can be overcome by requiring that A, B and C be defined upon the same domain, however, this is an artificial constraint and imposes a restriction that makes the logical system useless for application. For example, set $X = Y = Z = \mathbb{N}_{120}$ and interpret this as admissible ages, we can have statements like if (x is infant) then (x is young) which really does not tell us anything about age that we did not know from the constructions of the linguistic terms.

The aim of applied fuzzy logic is approximate reasoning an approximate reasoning is actually more ambitious in its goals than the creation of a fuzzy set extension of the modus ponens, modus tolens or any deduction scheme. In fuzzy set theory, as in reality, we don't really expect to have an exact match between the data observed and the fuzzy atomic variable of the premise. That is, data seldom fits the ideal model. This is a common situation in the real world, where a rule must be bent to fit a situation is was not designed to handle. With fuzzy set theory, implication is a very fluid notion especially when dealing with imprecise quantities.

Example 97. I am 52 years old and am in the fuzzy variable middle aged but only to a certain membership grade say 0.86. I am also young to a certain but lesser extent (0.02 by the formula of Eq. (16.1)). I am also a bit old (0.28 by Eq. (16.2)).

We don't want a scheme to deal with (hypothetical syllogism)

Premise Socrates is a man.

Inference rule All mean are mortal.

Deduction Socrates is a mortal.

We want a scheme to deal with

Premise This tomatoes is very red.

Inference rule If a tomato is red then the tomato is ripe.

Deduction The tomato is very ripe

or

$$\begin{array}{l} A' \\ A \to B \\ \hline B' \end{array}.$$

15.4. Fuzzy relations as logical representation

If we look at the result of any of the models of fuzzy implication, and especially at the tables in Example (93) we can notice that $I(a, b)$ is a fuzzy relation.

Often $A \to B$ is a fuzzy number. If A and B are fuzzy numbers and $\to$ is modeled by an implication operator that satisfies Axioms (1 and 2) and any of (5, 7 or 4) then $A \to B$ is a fuzzy number.

This is going to be important because it is time to generalize logical deduction schemes. Let us reconsider the generalized modus ponens:

$$\begin{array}{l} A' \\ A \to B \\ \hline B' \end{array} \tag{15.19}$$

To generalize the modus ponens we need a method to combine the fuzzy set A' and the fuzzy relation $A \to B$ to produce a fuzzy set B'. If we were to look through the earlier chapters of this book, there would be one outstanding candidate. Suppose A' has domain X and $A \to B$ has domain $X \times Y$. Then the sup–min composition of A' and $A \to B$ will produce a fuzzy set with domain Y, that is $A' \circ (A \to B)$ produces the exact kind of fuzzy set B' that we are looking to deduce. So the *fuzzy modus ponens* looks like

$$B' = A' \circ A \to B \tag{15.20}$$

Actually whenever we have a fuzzy set A' defined on X and a fuzzy relation R defined on $X \times Y$ then the sup–min composition can be considered as a conclusion B' about Y. This is the *generalized modus ponens* which looks like

$$B' = A' \circ R \tag{15.21}$$

Example 98. Fuzzy inference

Let $X = Y = \mathbb{N}_4$ and define the following fuzzy sets

Premise

x_1 is small—Fuzzy set representation $A = \frac{1.0}{1} + \frac{0.6}{2} + \frac{0.2}{3} + \frac{0.0}{4}$

Inference rule x_1 is approximately equal to x_2 or $x_1 \approx x_2$—Fuzzy set representation

$$E(x_1, x_2) = \begin{cases} 1 & \text{for } \langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle, \text{ and } \langle 4,4\rangle \\ 0.5 & \text{for } \langle 1,2\rangle, \langle 2,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle, \langle 3,4\rangle, \text{ and } \langle 4,3\rangle \\ 0 & \text{otherwise} \end{cases} \tag{15.22}$$

Deduction

$$\begin{aligned} A \circ E &= [1, 0.6, 0.2, 0] \circ \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 \\ 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.5 & 1 \end{bmatrix} \\ &= [1, 0.6, 0.5, 0.2] \end{aligned}$$

So that $B' = \frac{1.0}{1} + \frac{0.6}{2} + \frac{0.5}{3} + \frac{0.2}{4}$ which is similar to more or less small.

15.5. Selection of implication operators

15.5.1. Mathematical considerations

Of course a good question here is which implication operator I do we use to construct $A \to B$ from fuzzy sets A and B. many candidates were introduced in Chapter (3).

One of the major themes of fuzzy set theory is that the operations defined behave just like their counterparts in crisp set theory when the membership grades are restricted to $\{0, 1\}$ instead of $[0, 1]$. In Example (93) we have the fuzzy set $A = \frac{0.5}{1} + \frac{0.8}{2} + \frac{1.0}{3}$ and the fuzzy relation $R \equiv A \overset{z}{\to} B$, where $\overset{z}{\to}$ indicates Zadeh implication, given by the following table:

$A \xrightarrow{z} B$	1.0	0.8	0.2
0.3	0.7	0.7	0.7
0.8	0.8	0.8	0.2
1.0	1.0	0.8	0.2

This book has introduced sup–min composition in Chapter (8). Let us perform the sup–min composition of A and R which produces $A \circ R = \frac{1.0}{1} + \frac{0.8}{2} + \frac{0.5}{3}$. But $A \circ R$ is not equal to B and thus does not match the modus ponens rule.

What kind of composition satisfies the modus ponens, i.e., given fuzzy sets, A and B, and use an implication operator to construct the relation $R = A \rightarrow B$, what must composition look like so that we recover B: that is $B = A \circ R$. This question has been partially answered in a more general form. Chapter (8) introduces max–t composition performed with a t-norm t. The question rephrased is, given fuzzy sets A and B and an implication operator i to construct $R(x,y) = \mathsf{i}(A(x), B(y))$, what properties must i fulfill so that $B = A \overset{\mathsf{t}}{\circ} R$?

Definition 73. Let A be a normal fuzzy set. For any continuous t-norm t and the associated omega operator (also called the residuum, see Eq. (6.28)) ω_{t} and let $\mathsf{i} = \omega_{\mathsf{t}}$, that is, define $\mathsf{i}(a,b) = \omega_{\mathsf{t}}(a,b)$ then

$$B = A \overset{\mathsf{t}}{\circ} R$$

where $R(x,y) = \mathsf{i}(A(x), B(y))$. Equivalently

$$B(y) = \sup_{x \in X} \mathsf{t}\left[A(x), \mathsf{i}\left(A(x), B(y)\right)\right] \tag{15.23}$$

If the range of the membership function A for Equation (**??**) is $[0,1]$ then we have the following result.

Definition 74. The following fuzzy implication operators satisfy (**??**) if the range of A is $[0,1]$

1. Gaines-Rescher

2. Godel

3. Wu.

Name	standard intersection	algebraic product	bounded difference	drastic intersection
Early Zadeh	$\max\left[\frac{1}{2}, B\right]$	$\max\left[\frac{1}{4}, B\right]$	B	B
Lukasiewicz	$\frac{1}{2}(1+B)$	$\frac{1}{4}(1+B)^2$	B	B
Mamdani	B	B	B	B
Larsen	B	B	B	B
Standard Strict	B	B	B	B
Gödel	B	B	B	B
Gaines-Rescher	B	B	B	B
Goguen	$B^{\frac{1}{2}}$	B	B	B
Kleene-Dienes	$\max\left[\frac{1}{2}, B\right]$	$\max\left[\frac{1}{4}, B\right]$	B	B
Reichenbach	$\frac{1}{2-B}$	$\max\left[B, \frac{1}{4-4\min\left(B,\frac{1}{2}\right)}\right]$	B	B
Yager	B	B	B	B

Table 15.3.: Generalized modus ponens

Proof. Let us use i_{gr} then

$$\begin{aligned}
\sup_{x\in X} \mathsf{t}\left[A(x), \mathsf{i}_{GR}\left(A(x), B(y)\right)\right] &= \sup_{a\in[0,1]} \mathsf{t}\left[a, \mathsf{i}_{gr}\left(a, B(y)\right)\right] \\
&= \max\left\{\sup_{a\leq B(y)} \mathsf{t}\left[A(x), \mathsf{i}_{gr}\left(a, B(y)\right)\right],\right. \\
&\qquad \left.\sup_{a>B(y)} \mathsf{t}\left[A(x), \mathsf{i}_{gr}\left(a, B(y)\right)\right]\right\} \\
&= \max\left\{\sup_{a\leq B(y)} \mathsf{t}\left[A(x), 1\right], \sup_{a>B(y)} \mathsf{t}\left[A(x), 0\right]\right\} \\
&= \max\left\{\sup_{a\leq B(y)} A(x), \sup_{a>B(y)} 0\right\} \\
&= \max\left\{B(y), 0\right\} \\
&= B(y)
\end{aligned}$$

The proofs for the other two implication operators are similar.

□

The Table (15.3) shows the results of sup-t composition for four t-norms, standard intersection ($\min$), algebraic product (times), bounded difference, and drastic intersection.

The line of reasoning that produced Table (15.3) is that the generalized modus ponens, or the fuzzy modus ponens, must satisfy the traditional deduction scheme given in Eq. (3.4) which works out after some development to produce the requirement expressed in Eq. (**??**).

A similar development, based this time upon the deduction scheme for the modus tolens Eq.(3.6) would produce the requirement

$$B^{\mathsf{c}} = A^{\mathsf{c}} \overset{\mathsf{t}}{\circ} R$$

where $R(x, y) = \mathsf{i}(A(x), B(y))$ and c represents a complement operator. Equivalently we would have the expression

$$\mathsf{c}(B(y)) = \sup_{x \in X} \mathsf{t}\left[\mathsf{c}(A(x)), \mathsf{i}(A(x), B(y))\right] \tag{15.24}$$

as our requirement.

Finally developing the hypothetical syllogism, Eq. (3.8), would produce the requirement

$$R_{AC} = R_{AB} \overset{\mathsf{t}}{\circ} R_{BC}$$

where $R_{AC}(x, z) = \mathsf{i}(A(x), C(z))$, $R_{AB}(x, y) = \mathsf{i}(A(x), B(y))$, and $R_{BC}(y, z) = \mathsf{i}(B(y), C(z))$ This requirement is equivalently to the following expression

$$\mathsf{i}(A(x), C(z)) = \sup_{y \in X} \mathsf{t}\left[\mathsf{i}(A(x), B(y)), \mathsf{i}(B(y), C(z))\right] \tag{15.25}$$

Tables similar to Table (15.3) could then be developed that show which combinations of implication operators and t-norms satisfy Eqs. (15.24 and 15.25).

15.5.2. Heuristic considerations

Since there is no consensus on the proper implication operator in fuzzy set theory the decision of which operator to use is simply a matter of choice. There are two main branches of application to fuzzy logic, approximate reasoning (AR) and fuzzy control . There are endless papers on the mathematical properties of the various implication operators. There are also numerous papers comparing the results of various implication operators in fuzzy controllers, where they can be compared to ideal controllers (when an ideal controller exists).

However, data driven methods for the selection of implication operators in AR are more difficult to obtain. That is because AR deals with problems where there has never been suitable methods in conventional mathematics, other than Boolean logic, even for simple situations.

15.6. Fuzzy propositions

To understand logical modifiers we must first dissect what we have already constructed. In this chapter we have been using A and $A(x)$ and saying that they are fuzzy sets, and sometimes that they are linguistic terms, which are also fuzzy sets. In a logical statement like $A \to B$ the symbol A is called a propositions and $A(x)$ is a proposition about something (about x). So the fuzzy set A is really a fuzzy proposition about something (whatever the domain variable represents). The simplest type of fuzzy proposition p has as its prototype:

$$p : V \text{ is } G \tag{15.26}$$

where p is the proposition, V is a domain set and G is a fuzzy set on a variable v in the domain set V. This is the type of statements this book has been using all along. The most common example we have been using has V being the set $\mathbb{N}_{120} = \{1, 2, 3, ...120\}$ of ages and G is a linguistic term like young that is modeled with a fuzzy set defined on the set of ages. Sometimes we have been more specific and made a statement like

$$Mary \text{ is young.}$$

but as a proposition this really should be read

$$\text{age}(Mary) \text{ is young}$$

since it is the age of $Mary$ that has values in the set $\mathbb{N}_{120}$. The prototype of this type of statement looks like:

$$p : \mathcal{V}(i) \text{ is } G \tag{15.27}$$

where I is a set of individuals or objects having attributes in the set V modeled by fuzzy set G. The variable $\mathcal{V}$ is a function, $\mathcal{V} : I \to V$ for $i \in I$, that maps individuals to an attribute set. The prototype in Eq. (15.27) can be interpreted as

$$p : \text{age}(i) \text{ is young}$$

where i is the individual $Mary$.

15.7. Qualified proposition

Zadeh, and others also allow for statements that are not directly modeled in classical logic. The general form of this type of qualified prototype proposition is

$$p : V \text{ is } G \text{ is } Q\ . \tag{15.28}$$

An example of this type of model is the following statement

$$(x \text{ is } A) \text{ is true}$$

15.7.1. true

We will first examine a qualified proposition such as $(x \text{ is } A)$ is true exemplified as *Mary* is young is true

A proposition about an object A can be a statement about its truth. So truth must be a fuzzy concept. Subjectively we know that this is true. Traditional logic and and much of western philosophy has had at its heart a belief in a two valued logic, thus the reliance on dualism and the dialectic.

What is truth? Philosophers have argued about this endlessly. But as a linguistic term we need to first figure out the domain of linguistic terms like true. Since truth in logic takes on the numeric values 0 and 1 it makes sense that true should be a fuzzy set defined on $[0, 1]$ which evaluates evidence and says how we will weight that evidence as too its truthfulness. The idea here is that we need quite a bit of evidence about something before we say it is true. If a coin is flipped 100 times and comes up heads 51 times, there is not enough evidence to say that the coin, when flipped, will land heads up. On the other hand if a coin is flipped 100 times and comes up heads 81 times, there is enough evidence to say that the coin will land heads up. I would certainly like to be in Vegas betting with these odds.

Our intuition about truth says that it should be a fuzzy set $\mathbf{T}$ such that $\mathbf{T}(0) = 0$ and $\mathbf{T}(1) = 1$ Otherwise it is context dependent, depending about how we feel about the evidence, for example $\mathbf{T}_1(t) = t$, $\mathbf{T}_2(t) = t^2$, and $\mathbf{T}_3(t) = 1$ if $t = 1$ and 0 otherwise are all trues, in fact $\mathbf{T}_2 =$very$(\mathbf{T}_1)$. So the fuzzy set for a statement like x is small is very true or "x is A is $\mathbf{T}_2$" which is evaluated as just $A \circ \mathbf{T}_2(x)$, the (functional) composition of the functions A and $\mathbf{T}_2$. If $A = \frac{1.0}{1} + \frac{0.6}{2} + \frac{0.2}{3}$ (or $A = \{\langle 1, 1\rangle, \langle 2, 0.6\rangle, \langle 3, 0.2\rangle\}$) then x is small is very true is the fuzzy set

$$A \circ \mathbf{T}_2 = \frac{1.0}{1} + \frac{0.36}{2} + \frac{0.04}{3} \tag{15.29}$$

since the function $\mathbf{T}_2(t) = t^2$ just takes a membership grade and squares it.

The logical constant false is **not** the complement of true, instead it is the mirror image. That is, once we decide on a fuzzy set function $\mathbf{T} : [0,1] \to [0,1]$ to model true then the function $\mathbf{F}$ to model false is defined as its mirror image:

$$\mathbf{F}(t) = \mathbf{T}(1-t)$$

for $t \in [0,1]$.

15.7.2. Quantifiers

The two quantifiers used in classical logic are $\forall$ (for all) and $\exists$ (there exists). For example the statement $\exists x \in X\ x < 1$ is read that there exists an element x of X that is less than one. Note that this is true if X is the integers but is false if X is the natural numbers. In approximate reasoning we generalize this completely to a quantifier Q that makes a statement about the number of objects x that satisfy the proposition. For example consider the statement that "this class has five young students". This is equivalent to the stilted English statements "students that are young number five" or "students is young is five" which is of the form given in Eq. (15.28).

The techniques available to us allow for the conclusion of a single truth vale to the proposition. This is the situation where the class has already started and data about the students ages are available. In this case the age of each student can be input to the function that models the fuzzy set young. Finally the sigma count of this fuzzy set provides an approximate number of young students. This value is then plugged into the fuzzy set for five to produce a final truth value.

Example 99. Let the data in the following table represent the ages of the students in the class ACT101 be given in the Table (15.4) and assume that young is the fuzzy set of Eq. (16.1) of Chapter (16). Further assume that five is modeled with the triangular fuzzy set $T_{tri}\langle 3, 5, 7\rangle$. The third column of the table gives the membership grade of the age variable in the fuzzy set young. The sigma count of young is then

$$\begin{aligned} \|young\| &= 1.00 + 0.73 + 0.12 + 0.28 + 0.05 + 1.00 + 1.00 + 0.50 + 0.04 + 1.00 \\ &= 5.72 \end{aligned}$$

Name	**Age**	young
Ann	21	1.00
Bob	28	0.73
Cat	38	0.12
Dee	33	0.28
Eli	46	0.05
Fan	24	1.00
Gil	22	1.00
Han	30	0.50
Ira	50	0.04
Jal	24	1.00

Table 15.4.: The ages of the students in the class ACT101.

and the membership grade of 5.72 in the triangular fuzzy set five $= T_{tri}\langle 3, 5, 7\rangle$ is 0.64 and this is the truth vale of the statement "this class has five young students".

Other situations, such as classes undergoing enrollment, with a partial class list along with historical data on enrollment and age data allow only a fuzzy conclusion as to the truth of the statement "this class has five young students." Usually this situation is dealt with by examining the statement "this class has n young students" and constructing a possibility distribution on n.

15.7.3. Putting the pieces together

Approximate reasoning is applied fuzzy logic. It involves many pieces that have already been introduces in the previous chapters of this book. The first piece is a linguistic system (Ch. 16) which uses fuzzy numbers (Ch. 7) to model terms such as young and old. The second piece is fuzzy logic which allows for the constructions of fuzzy sets from its component pieces, this allows us to model connectives such as and and or. Fuzzy logic also allows the representation of logical statements such as if $Pressure$ big then $Temperature$ high. Then $\sup$-$\min$ composition allows for all the pieces to be put together to produce conclusions.

15.8. Possibility theory

Much of the literature of fuzzy logic is cast in the language of possibility theory since possibility theory requires that $r(x_1) = 1$. However fuzzy logic often uses unsorted possibility distributions on continuous domains so that $r(x_i) = 1$ for some x_i. All of the fuzzy numbers and relations in this chapter have always had some element of membership grade 1.

15.9. Notes

Much of this Chapter is pure Zadeh. His paper in *Fuzzy logic for the management of uncertainty* Zadeh and Kacprzyk (1992) covers much of the material of this chapter and other papers in this book give an excellent overview of the application and research areas in this field. An early collection of his papers by Yager et al. (1987a) details many of his contributions.

Also relevant are the books *Fuzzy Sets, Decision Making and Expert Systems Zimmermann (1987), An Introduction to Fuzzy Logic Applications in Intelligent Systems* Yager and Zadeh (1994), *Fuzzy Logic A Practical Approach* McNeill and Thro (1994) and *The Fuzzy Systems Handbook* Cox (1994b).

Approximate reasoning based on fuzzy predicate logic was studied by many authors besides Zadeh, including Gaines and Shaw (1986), Baldwin (1979), Baldwin and Guild (1980), Baldwin and Pilsworth (1980), Turksen and Yao (1984), and Goguen (1979).

Elkand published an infamous paper, proving that under suitable axioms, the law of the excluded middle had to be true, thus negating fuzzy set theory. Of course this ignores the fact that logic only shows that axioms entail deductions. Unfortunately, Elkand's axioms are not even true for Łukasiewicz 3-valued logic. Łukasiewicz (1963) demonstrates that 3-valued logic is a superset of 2-valued logic.

15.10. Homework

Let us define the universal sets $X = \{1, 2, 3\}$ and $Y = \{a, b, c\}$ and the fuzzy sets D, E, F, G, and H.

$$D(x) = \frac{0.5}{1} + \frac{0.5}{2} + \frac{1.0}{3} \tag{15.30}$$

$$E(x) = \frac{0.2}{1} + \frac{0.6}{2} + \frac{0.4}{3} \tag{15.31}$$

$$F(y) = \frac{0.3}{a} + \frac{0.7}{b} + \frac{0.9}{c}$$

$$G(x) = \frac{1}{x} \tag{15.32}$$

$$H(y) = \frac{ascii(y) - 96}{3}$$

where $ascii(y)$ is the numerical value of the character x in the ascii table.

1. What is the implication operator generated by $f(x) = \ln(1 + \lambda x)$ and what is its corresponding complement operator. Name the complement operator.

2. What is the implication operator generated by $f(x) = \frac{2x}{1+x}$ and what is its corresponding complement operator. Name the complement operator.

3. What is the implication operator generated by $f(x) = x^2$ and what is its corresponding complement operator. Name the complement operator.

4. Compare the results of the three questions above, especially the respective complement operators. What does this tell you.

5. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Gougen's implication operator? In other words, fill out the following Table where $\stackrel{gg}{\to}$ indicates that implication is done with i_{gg}.

$E \stackrel{gg}{\to} F$	a	b	c
1			
2			
3			

6. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Lukasiewicz's implication operator?

$E \stackrel{a}{\to} F$	a	b	c
1			
2			
3			

7. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Larsens's implication operator?

$E \stackrel{l}{\to} F$	a	b	c
1			
2			
3			

8. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Godels's implication operator?

$E \stackrel{g}{\to} F$	a	b	c
1			
2			
3			

9. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Mamdani's implication operator?

$E \stackrel{mm}{\to} F$	a	b	c
1			
2			
3			

10. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with the Standard Strict implication operator?

$E \stackrel{ss}{\to} F$	a	b	c
1			
2			
3			

11. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Reichenbach's implication operator?

$E \xrightarrow{r} F$	a	b	c
1			
2			
3			

12. Given the fuzzy sets E and F (Eq. (15.31)) what is the truth value of $E \to F$ for each of the nine elements of $X \times Y$ if we model implication with Yager's implication operator?

$E \xrightarrow{y} F$	a	b	c
1			
2			
3			

13. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \to H$ for each of the nine elements of $X \times Y$ if we model implication with Gougen's implication operator? In other words, fill out the following Table where $\to_{gg}$ indicates that implication is done with ¡$_{gg}$.

$G \xrightarrow{gg} H$	a	b	c
1			
2			
3			

14. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \to H$ for each of the nine elements of $X \times Y$ if we model implication with Lukasiewicz's implication operator?

$G \xrightarrow{a} H$	a	b	c
1			
2			
3			

15. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \to H$ for each of the nine elements of $X \times Y$ if we model implication with Larsens's implication operator?

$G \xrightarrow{l} H$	a	b	c
1			
2			
3			

16. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \to H$ for each of the nine elements of $X \times Y$ if we model implication with Godels's implication operator?

$G \xrightarrow{g} H$	a	b	c
1			
2			
3			

17. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \rightarrow H$ for each of the nine elements of $X \times Y$ if we model implication with Mamdani's implication operator?

$G \overset{mm}{\rightarrow} H$	a	b	c
1			
2			
3			

18. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \rightarrow H$ for each of the nine elements of $X \times Y$ if we model implication with the Standard Strict implication operator?

$G \overset{ss}{\rightarrow} H$	a	b	c
1			
2			
3			

19. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \rightarrow H$ for each of the nine elements of $X \times Y$ if we model implication with Reichenbach's implication operator?

$G \overset{r}{\rightarrow} H$	a	b	c
1			
2			
3			

20. Given the fuzzy sets G and H (Eq. (15.32)) what is the truth value of $G \rightarrow H$ for each of the nine elements of $X \times Y$ if we model implication with Yager's implication operator?

$G \overset{y}{\rightarrow} H$	a	b	c
1			
2			
3			

21. Given the fuzzy sets D, E, and F (Eqs. (15.30 and 15.31) what is the result of $D \circ E \rightarrow F$ if we model implication with Gougen's implication operator? Remember that $\sup - \min$ composition is performed like matrix multiplication and you should have the tables for $E \rightarrow_{gg} F$ from the previous chapters homework.

22. What is the result of $D \circ E \rightarrow F$ if we model implication with Lukasiewicz's implication operator?

23. What is the result of $D \circ E \rightarrow F$ if we model implication with Larsens's implication operator?

24. What is the result of $D \circ E \rightarrow F$ if we model implication with with Godels's implication operator?

25. What is the result of $D \circ E \rightarrow F$ if we model implication with Mamdani's implication operator?

26. What is the result of $D \circ E \rightarrow F$ if we model implication with the Standard Strict implication operator?

27. What is the result of $D \circ E \rightarrow F$ if we model implication with Reichenbach's implication operator?

28. What is the result of $D \circ E \rightarrow F$ if we model implication with Yager's implication operator?

29. Given the fuzzy sets D, G, and H (Eqs. (15.30 and 15.32) what is the result of $D \circ G \rightarrow H$ if we model implication with Gougen's implication operator? Remember that $\sup - \min$ composition is performed like matrix multiplication and you should have the tables for $G \rightarrow_{gg} H$ from the previous chapters homework.

30. What is the result of $D \circ G \rightarrow H$ if we model implication with Lukasiewicz's implication operator?

31. What is the result of $D \circ G \rightarrow H$ if we model implication with Larsens's implication operator?

32. What is the result of $D \circ G \rightarrow H$ if we model implication with with Godels's implication operator?

33. What is the result of $D \circ G \rightarrow H$ if we model implication with Mamdani's implication operator?

34. What is the result of $D \circ G \rightarrow H$ if we model implication with the Standard Strict implication operator?

35. What is the result of $D \circ G \rightarrow H$ if we model implication with Reichenbach's implication operator?

36. What is the result of $D \circ G \rightarrow H$ if we model implication with Yager's implication operator?

37. Which of the above results seems most reasonable?

16. Linguistic variables

16.1. Complexity -vs- Precision

If a concept is sufficiently simplified we can measure it with precision. For example, my weight is 155 lbs. If it is artificial we can count it exactly. For example my wallet contains $23.

The real world is complex. How do I measure my health? What about my wealth? Do I know the value of everything I own at this moment. How do we measure well–being?

The real is hard to count. How many stars are there in the heavens? Is your night vision better or worse than mine? Is it cloudy out? You have binoculars! Is that cheating?

Many of the great philosophers have argued this problem under various guises. Ontology is the study of knowledge. Poetics is the study of symbols?

How can we have absolute knowledge? It is very difficult when you think about it. For example my weight is really 154 lbs. 13 oz. and I am sure there are devices that can calculate my weight with much greater precision than what is given by the figure "154 lbs. 13 oz".

Bertram Russell came right out and said that symbolic logic has nothing to do with our everyday existence. Mathematics is also meaningless. Computers simply manipulate symbols in a preprogrammed paths.

Cybernetics

Cybernetics, which encompasses and is sometimes studied under the names of

- Adaptive systems
- Artificial intelligence
- Artificial life
- Chaos theory
- Decision science
- Information science
- Man-machine studies
- Neural networks
- Operations research
- Optimization methods

- Symbiotics
- Systems science

deals with the problems of understanding human beings and how they interfacing with machines. It especially has come to mean; how human beings interact with computer technology.

Ergonomics is the study of designing machines to fit humanity. A good example of this is the can opener. Twenty years ago, the can opener was made of two thin arms of metal, a fixed cutting blade mounted on the top arm and a gear that rotated the can. It was difficult to grasp, often caused spills and the blade could easily contaminate the food. A modern can opener has two wide, rubber-padded handles and a rotating cutting blade. It doesn't slip or jam. The can seldom falls onto the floor. This is good ergonomics. It shows great understanding of the mechanics of can opening and the mechanics of the human hand, and the correct way that these two systems interact. A modern hammer might contain a tuning fork to dampen the vibrations that are transmitted to the hand and wrist of the hammerer.

Unfortunately the human hand, and the hammer are both much easier to understand than the human brain and the computer. We are not very good at designing technology that fits our brain. Computer programming is difficult because it is unnatural. Logic is unnatural, humans rationalize, not reason.

Human don't understand much. Who would like to try and completely explain the workings of the can opener. How much torque is involved. Where does the plastic come from? Who makes these can openers. Are there left handed versions available? What kind of metal is the cutting wheel made of?

Few people would like to get up in front of class and categorically defend their knowledge on any subject. Not me. Not about anything!

We don't understand how a cat works. We don't understand how we work. We especially don't understand how our brains work. Maybe this is why psychology is one of the most popular majors for college students. Most of us don't understand how our car, computer, microwave, etc. works. Even those who understand how a computer works don't really understand it. It depends on quantum physics which is statistical and only deals in probabilities not absolute knowledge.

Traditional areas for cybernetic problems include

- Artificial intelligence—making the computer work like the brain.
- Control—making machines react to their environment.
- Heuristics—translating the rules of thumb that humans use into something machines can understand.
- Pattern Recognition—making machines learn from example, the way human's learn
- Prediction—getting computers to extrapolate trends in the data.
- Risk analysis—getting machines to dealing with contradictory problems.
- Syzygy—gestalts, the problem of parts and wholes. Making big problems into a bunch of little problems that we can solve and reintegrating the results.

These fields have gained many adherents because of the failure of traditional methods to tackle the problems posed by large and complex systems. This dilemma has been exacerbated by the rise of computer technology.

Here is a simple example. A weather researcher runs a program to simulate convection currents in the air. The results are interesting so he desires to rerun a short segment of the simulation. Since the simulations takes a lot of time and computer power he feeds in the data vales of the system at the time he is interested in studying off of a printout. When he runs the simulation it is completely different from the original simulation. The reason for this disparity is that the computer printout had the data to eight decimal places while the computer had much greater internal precision. But there is even more to the story. The researcher then experimented with how the precision of the data representation altered the weather forecast. A C/C++ program could use float or double as the data type of its variables. Two programs, whose only difference is the precision of the data type, give very different forecasts for the weather after the first few hours. This was the beginning of the study of Chaos. This is why the weather man can never predict the weather next week. To do this he would have to have a computer with infinite precision, and he would have to measure all the data from the real world with infinite precision. Wherever he chops off a measurement, 10 decimal places, 20 decimal places, etc., will change the long range weather forecast.

A traditional solution method for a complex problem would involve the following steps:

1. Define a mathematical model.
2. Gather data.
3. Simulate/solve the model to the required precision.
4. Analyze the results.

Sometimes steps one and two are reversed. The data are gathered and the data suggests a proper mathematical model. However the data often can be used to fit more than on mathematical model. Anyone who has studied algorithms in computer science knows that there is more than one way to accomplish many goals, such as sorting the data. The size and type of the data may make one particular type of sorting program more efficient in application. For example quicksort is, on average, one of the fastest sorting methods for real numbers, but when quicksort is slow, it can be very slow. Heapsort is, on average, very slightly slower than quicksort, however, the heapsort always runs at exactly the same speed, so it is very reliable.

The problems with the traditional solution method for complex problems include:

1. No mathematical model fits the data or many mathematical models fit and they give contradictory results.
2. The mathematical model is difficult to simulate with present or foreseeable computer limitations.
3. The mathematical model is impossible to solve.
4. Data is difficult to get, contradictory, vague or just plain nonexistent.

5. Chaos is everywhere in complex systems, predictability will then depend on measurement precision.
6. The pieces don't fit together (this is what syzygy deals with).

16.2. The black box approach

There are a lot of different methodologies kicking around at the beginning of the 21st century that attempt to deal with some or all of the problems with traditional methods of solving complex and/or vague problems. These methodologies include:

1. Genetic algorithms
2. Neural networks
3. Simulated annealing
4. Random search
5. Fuzzy sets
6. Rough sets
7. Probability
8. Random sets

One thing many of these fields have in common is the black box approach. They don't claim to explain the phenomena, just to be able to make deductions or conclusions about them. Their attitude is; given enough I/O data, can we build a simulation model without understanding anything about how the system works? At first this may seem unreasonable, yet if we think about it, our brains have some kind of simulation of a car inside them that works sufficiently well to allow us to drive home, without any precise measurement data inside our brains, and this does not worry us. In fact biological methods are the inspiration of many of the modern cybernetics approaches. Thus cybernetics is a two way street, it use biologically inspired methods and one of its major goals is the explanation of biological experiments. The brain is trying to understand itself.

Statistical methods were the traditional approach to solve problems in the face of uncertainty. Uncertainty was translated as randomness by the researchers who developed probability. This method does work if the uncertainty in the system is sufficiently regular, i.e., random. One could also try to model it as a finite state machine or transition diagram.

However, not everything that is uncertain seems to fit the probabilistic mold. Suppose someone asks you, "How did you like that restaurant?" Now you liked parts of it and disliked parts of it, so there in uncertainty in your mind. You might say "I liked it 50%" but this is unlikely. It is more likely you will say something like "It was OK." This is not a probabilistic statement. It is an imprecise statement.

What is needed to deal with complex and imprecise systems is a methodology that can augment or process vague and incomplete input data to get a desirable output? In most cases it is impossible to get an exact match between the results of our

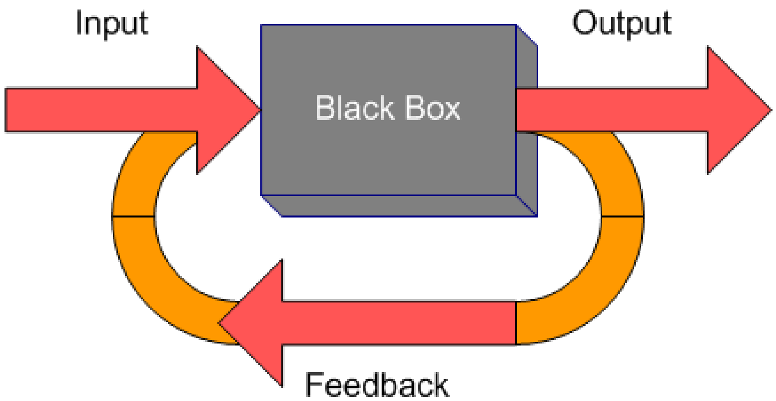


Figure 16.1.: Black box.

methodology and the real world situation. Output may only be able to steer us in the right direction. Is an "OK" restaurant good enough for tonight's meal if it is close? How about if the restaurant is close and inexpensive?

How do humans deal with the complex, vague, and incomplete data which the real world throws at us in ever larger and faster gulps? We use language. We use abstraction, simplification and symbols to reason about complex things. The symbols are usually words and in human language words are not precise, at least not very often. "two" is a precise word, however it refers to an idealized abstract entity. "tree" is an imprecise word, it refers to an abstraction of the real world that is not precisely defined. The word "tree" includes palm trees, pine trees, and peach trees, which have some similarities and some differences.

Philosophers argue these problems using big words like ontology, epistemology, and "the thing in itself" as if the thing were ever anything else. Of course the thing is something else, its what we perceive it is, which is not necessarily precise and objective. Anybody who has ever fallen in love or trusted someone else has learned that there are many perceptions that in long run do not match reality.

It is at this point in a philosophy class that everyone starts to argue about what the words mean. The answer to what the words means is: "It means exactly what I said." except that I do not always know what I am saying.

young

The word young is a linguistic term and like most linguistic term it is not precise.

This is common to all languages. They evolve. Words change their meaning over time. Awful meant full of awe or awesome.

Slang changes very fast. Look at the word rap. cool means acceptable. slime means unprincipled. In slang bad means very good.

As humans we understand the word cat without ever having seen the collection of all cats in the world.

We have generalized from a collection of specific instances a meaning.

Depending on the time and place cat may include domesticated and wild cats. It may mean just the common house cat or it may include all felines.

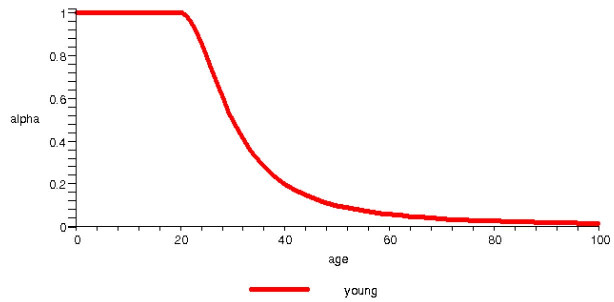


Figure 16.2.: A graph of the membership function of the linguistic variable young.

We use cat without knowing the anatomy, absolute dimensions or defining characteristics of cat.

Even if we were the worlds expert in cat the human races ignorance of biology and neurobiology would leave us sorely lacking in a complete understanding of cat.

This bothers us not at all. We all use the word cat to mean what we want it to mean.

Most everyone has an opinion about everything and most people understand absolutely nothing, and work hard to keep it that way.

16.2.1. Fuzzy sets and natural language

Human language is a powerful tool. Some think it was one of the two basic tools that caused the evolution of modern humanity (along with hands). Language is instinctive in human behavior, a child does not have to be taught a language, they will absorb whatever language they are in contact with, whether it is English or Mandarin. If many languages are spoken, such as in a port city, the children inevitably evolve a polyglot language, a creole. However, before fuzzy set theory, the only useful mathematical models for human language was logical and logic is severely limited in application. Fuzzy set theory however is uniquely applicable in the study and manipulation of language based systems, linguistic systems.

Let the universe U be the positive real numbers, $U = \mathbb{R}^+$. Define a fuzzy set Y with membership grade given by

$$Y(u) = \begin{cases} 1.0 & 0 \leq u \leq 20 \\ \left[1 + \left(\frac{u-20}{10}\right)^2\right]^{-1} & u > 20 \end{cases} \tag{16.1}$$

for any $u \in U$ The fuzzy set Y is a decreasing s–shaped fuzzy set.

Remark 10. $Y(u)$ is s-shaped fuzzy number if we abandon the bounded support requirement of fuzzy numbers.

The graph of the function Y looks like:

If we think of $u \in U$ as an age then Y is the fuzzy set that represents the linguistic variable young.

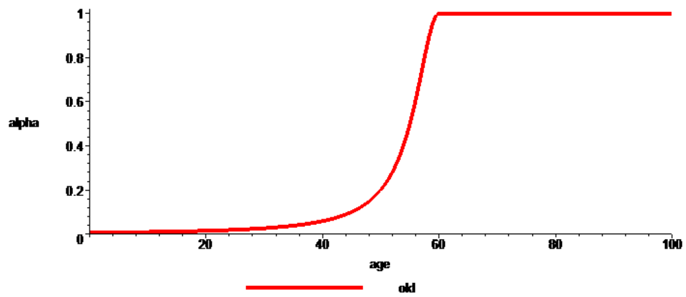


Figure 16.3.: A graph of the membership function of the linguistic variable old.

Define a fuzzy set O (which is an increasing s–shaped fuzzy set)

$$O(u) = \begin{cases} \left[1 + \left(\frac{u-60}{10}\right)^2\right]^{-1} & 0 < u \leq 60 \\ 1.0 & u > 60 \end{cases} \tag{16.2}$$

The graph of this function indicates that O is the fuzzy set that represents the linguistic variable old.

16.3. Linguistic System

Both these fuzzy sets Y–young and O–old are instances of term's whose subject is age. Any fuzzy set defined upon the universe U of ages is a an exemplification or a statement about a group of ages. All of this theory was worked out by Zadeh many years ago Bellman and Zadeh (1970), Zadeh (1973). In Zadeh's work a linguistic variable is the quintuple $\langle \mathcal{X}, T(\mathcal{X}), U, G, M \rangle$ where:

- $\mathcal{X}$—Age, a linguistic variable,
- U— The numerical ages from 0 to ∞, the universal set,
- T—young, adolescent, infant, old, etc., Terms associated with $\mathcal{X}$,
- G—The rules for generating a phrases or name about age, $\mathcal{X}$, like very old, somewhat old, etc., a grammar,
- M—A method of generating a fuzzy set $M(\mathcal{X})$ for the terms in T and for all expressions generated by the grammar G. For example, the term young on the linguistic variable $\mathcal{X}$ is assigned the fuzzy set Y, whose membership function (also called its compatibility) is given in Eq. (16.1).

The difference between U and $\mathcal{X}$ is that U is just a set of numbers, and could represent temperature in a different linguistic system. The linguistic set $\mathcal{X}$ tells us how the values in U are interpreted. In this linguistic system the numbers in U are interpreted as ages in years. In a different linguistic system the numbers in U could be interpreted as temperature in degrees Kelvin.

essentially	virtually	very
sort of	rather	not
technically	almost	much
kind of	regular	fairly
actually	mostly	pretty
loosely speaking	in essence	barely
strictly	basically	reasonably
roughly	principally	extremely
in a sense	lower than	indeed
relatively	higher than	really
practically	particularly	more or less
somewhat	largely	pseudo-
exceptionally	for the most part	nominally
anything but	strictly speaking	literally
often	especially	typically
damn		

Table 16.1.: Linguistic Hedges.

16.3.1. Hedges

Already introduced were words like very or somewhat that alter meaning. How about phrases like young to middle aged. Zadeh calls them hedges Zadeh (1972a). Like when we hedge a bet.

Already introduced were words like very or somewhat that alter meaning.

How about phrases like young to middle aged.

Zadeh calls them hedges. Like when we hedge a bet. In the English language they are part of the grammar that helps us be more precise about the age (or other subject) that we are discussing. These hedges need to be added to the grammar G. Table 16.1 contains a list of some of the common hedges of the English language.

How do we represent hedges?

Is there any similarity between the transformation from young to very young and the transformation from old to very old. I would certainly think so and so did Zadeh. He defined hedges in terms of operators on the fuzzy set membership functions. Here is a short list of some common operators:

CONCENTRATION—$con(A)(u) = A^2(u)$

DILATION—$dil(A)(u) = A^{1/2}(u)$

NORMALIZATION—$norm(A)(u) = \frac{A(u)}{\max_{u \in U} A(u)}$

INTENSIFICATION—$int(A)(u) = \begin{cases} A^2(u) & A(u) < 0.5 \\ A^{1/2}(u) & A(u) \geq 0.5 \end{cases}$

Traditionally

very $A = con(A)$

more or less $A = dil(A)$

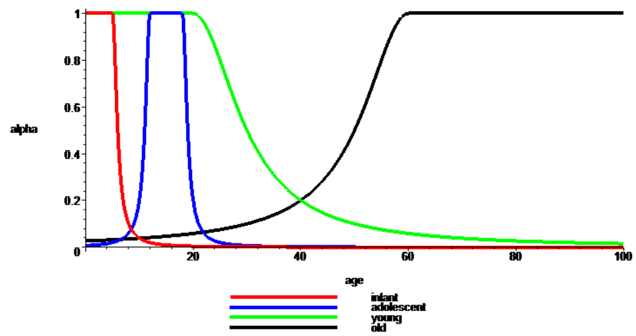


Figure 16.4.: Graphs of the membership functions of some linguistic terms.

plus $A = A^{1.25}$

slightly $A = int[plus(A)$ and not $very(A)]$

Besides the hedging operators we are using the fuzzy set connective and or intersection and not or complement. As we saw in the previous chapter these can be modeled with any t-norm and any complement operator. All of the hedge operations, con, dil, int, and $plus$ map a membership grade of one to one, i.e. $con(1) = 1$, $dil(1) = 1$, $int(1) = 1$, $plus(1) = 1$.

Example 100. The fuzzy set for the term young is given by Eq. (16.1). The equation for the term very young would be a concentration of the equation for the term young. Hence its equation would be $con(Y)(u) = Y^2(u)$ or

$$Y^2(u) = \begin{cases} 1.0 & 0 \le u \le 25 \\ \left[1 + \left(\frac{u-25}{5}\right)^2\right]^{-2} & u > 25 \end{cases} \tag{16.3}$$

16.3.2. Grammar

If you are have ever had to learn a programming language then you already have some experience with a formal grammar. When learning C you are first introduced to simple arithmetic statements like

```
x=x+10;
```
(16.4)

then to conditional statements such as

```
if(i<5) x=x+10;
```
(16.5)

and finally to repetition in the form of

```
for(i=0;i<10;i++) x=x+10;
```
(16.6)

Then these pieces; statements, conditions, and loops, are combined in many ways to produce complex functionality. The rules of combination are the rules of the grammar of C/C++. For example `if(i<5) x=x+10;` has as one of its components the simple statement `x=x+10;`

The most common way to write a grammar is something called Banach Normal Form..The grammar is defined recursively. This means that it is specified by rules of construction that can be applies in any combination and in any order.

Fir the linguistic grammars we are interested in here the basic building blocks are terms defined on linguistic variables. The symbol "::=" translates as "is defined to be" and the symbol "|" separates two exclusive choices. Here is an example of a grammar **G1** designed to handle the language of the ages.

<Primary Term> ::=young | old | infant | adolescent

<Hedge> ::= very | slightly | more or less | plus

<Range Phrase> ::= <Hedged Primary> to <Hedged Primary>

<Hedged Primary>::= <Hedge><Primary> | <Primary>

Example 101. The word young is a primary term and the word very is a hedge hence the rule <Hedged Primary>::= <Hedge><Primary> allows us to construct the prase very young in our grammar.

16.3.3. Method

We have all the pieces now to precisely define the linguistic system $\langle \mathcal{X}, T(\mathcal{X}), U, G, M \rangle$.

- $\mathcal{X}$ is Age.
- U is the universal set $\{0, 1, 2, 3, ..., 99, 100\}$.
- T is the set $\{\mathsf{young}, \mathsf{old}, \mathsf{infant}, \mathsf{adolescent}\}$
- G is the grammar
  <Primary Term> ::=young | old | infant | adolescent

<Hedge> ::= very | slightly | more or less | plus

<Range Phrase> ::= <Hedged Primary> to <Hedged Primary>

<Hedged Primary>::= <Hedge><Primary> | <Primary>

- M consists of the fuzzy sets associated with the primary terms

$$M(\mathsf{young})(u) = Y(u) \tag{16.7}$$
$$M(\mathsf{old})(u) = O(u) \tag{16.8}$$
$$M(\mathsf{infant})(u) = I(u) \tag{16.9}$$
$$M(\mathsf{adolescent})(u) = A(u) \ . \tag{16.10}$$

These primary terms have the explicit fuzzy sets:

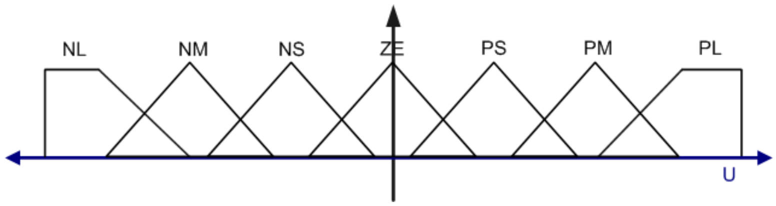


Figure 16.5.: Default granulation of a domain.

$$Y(u) = \begin{cases} 1.0 & 0 \le u \le 20 \\ \left[1 + \left(\frac{u-20}{10}\right)^2\right]^{-1} & u > 20 \end{cases} \tag{16.11}$$

$$O(u) = \begin{cases} \left[1 + \left(\frac{u-60}{10}\right)^2\right]^{-1} & 0 < u \le 60 \\ 1.0 & u > 60 \end{cases} \tag{16.12}$$

$$I(u) = \begin{cases} 1.0 & 0 \le u \le 5 \\ \left[1 + (u-5)^2\right]^{-1} & u > 5 \end{cases} \tag{16.13}$$

$$A(u) = \begin{cases} \left[1 + (u-12)^2\right]^{-1} & 0 < u \le 12 \\ 1.0 & 12 < u < 18 \\ \left[1 + (u-18)^2\right]^{-1} & u \ge 18 \end{cases} \tag{16.14}$$

The grammar rule <Hedged Primary>::= <Hedge><Primary> | <Primary> involves two cases. In the case <Hedged Primary>::= <Primary> then nothing is done to the mapped function. In the case <Hedged Primary>::= <Hedge><Primary> the mapping follows the rules Traditionally

very $A = con(A)$

more or less $A = dil(A)$

plus $A = A^{1.25}$

slightly $A = int[plus(A)$ and not $very(A)]$

Finally the grammar rule <Range Phrase> ::= <Hedged Primary> to <Hedged Primary> is handled by taking the union of the fuzzy sets so that very young to more or less young would be the union of the fuzzy sets $con(Y) = Y^2$ and $dil(Y) = Y^{1/2}$.

16.3.4. Granularity

Fig. (16.6) is a graph of both the young function Y and old function O on a single graph. If we are talking about age and only have the terms young and old we see that a lot of ages that are not included in these two categories to any great extent; the middle ages from 20 to 30. If age is going to be the topic of discussion then our vocabulary should have terms that parcel up the age domain in overlapping pieces. We need to introduce the fuzzy number middle as seen in Fig. (16.7).

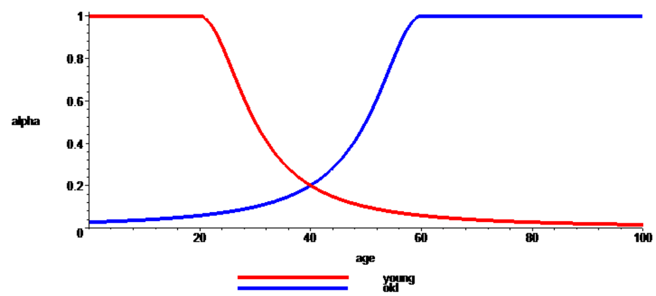


Figure 16.6.: The linguistic terms young and old.

Defining fuzzy subsets over the domain of a variable is referred to as granulation, granularity, or variable granulation, – in contrast to the division of a domain set into crisp subsets which is called quantization.

- Granulation results in the grouping of objects into imprecise clusters of fuzzy granules,
- The objects forming a granule are drawn together by similarity,
- Granulation can be seen as a form of fuzzy data compression,

Often granulation is obtained manually through expert interviews. If expert knowledge on a domain is not available, an automatic granulation approach can be used. Usually the domain is divided into an odd number of granules. Often the names are generic such as; NL: negative large, NM: negative medium, NS: negative small, Z: zero, PS: positive small, PM: positive medium, and PL: positive large. If we divide age into three groups we would have PS for positive small age, PM for positive medium age, and PL for positive large age. This section uses the subjective and more descriptive terms young $= PS$, middle $= MP$, and old $= LP$. Fig (16.5) shows an example of default granulation on a domain.

The design and construction of fuzzy sets is the topic of Chapter 9. For now we will assume that the fuzzy numbers represent our subjective opinion of the correct shape.

Using fuzzy sets allows us to incorporate the fact that no sharp boundaries exist between age groups. The fuzzy sets for old and young overlap to a certain extent.

In the Fig. (16.4) the granulation is fine at the lower end of the age spectrum where infant, and adolescent subdivide young and course at the upper end where everyone who is not young is just old. This might be the way a teenager sees the world, so granularity is strongly problem dependant. Both Fig. (16.4) and Fig. (16.7) are different granulations of the same domain set U, which further illustrates the difference between U, the numbers, and $\mathcal{X}$, age, viewed linguistically.

16.4. Notes

Linguistic systems will for the foundation of Approximate Reasoning (Chapter 15.3) and Fuzzy Control (Chapter 17).

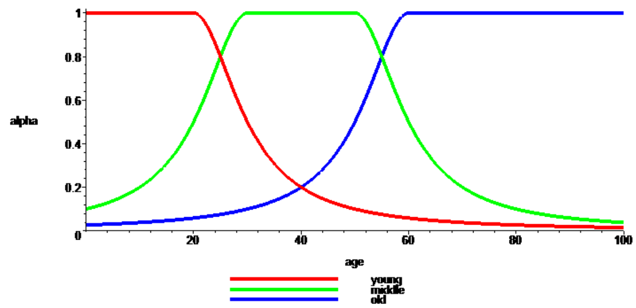


Figure 16.7.: The linguistic terms young, middle, and old.

Zadeh writes excellent explanations for both the theory and application of linguistic systems in Zadeh (1987a), and Zadeh (1987b). The book *Computing with words* Wang (2001) is a current resource in the applications of linguistic systems. Other applications are in Wenstop (1980), Oden (1984).

The major book on possibility theory is Dubois and Prade (1988).

16.5. Homework

Let H (in feet) for height be

$$H = \{0, 1, 2, 3, 4, 5, 6, 7, 8.9.10\}\,, \tag{16.15}$$

and let B (in number of burritos) be

$$B = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}\,. \tag{16.16}$$

1. Nanotechnology is a physical science, the construction of very tiny devices. Fuzzy sets are a non-physical technology. Search for a new non-physical technology that is not on the list at the beginning of this chapter. Describe the technology and its application.

2. What are some appropriate hedges when talking about height, and what are the appropriate operators to represent these hedges

3. Specify the grammar for talking about height.

4. Use all the rules in your grammar to generate a complex statement and show its mapped fuzzy set. Graph the mapped fuzzy set.

5. Define the fuzzy sets F for famished and H for HUNGRY on B.

6. What are some appropriate hedges when talking about burritos, and what are the appropriate operators to represent these hedges

7. Specify the grammar for talking about burritos.

8. Use all the rules in your grammar to generate a complex statement and show its mapped fuzzy set. Graph the mapped fuzzy set.
9. What is your major. In your major what are the important topics of discussion? Are any of these topics amenable to the construction of a Zadeh type fuzzy grammar linguistic systems? How would you go about this.
10. What is your hobby. In your hobby what are the important topics of discussion? Are any of these topics amenable to the construction of a Zadeh type fuzzy grammar linguistic systems? How would you go about this.
11. Critique Zadeh's linguistic system. Would you have done something different or even labeled things differently.

17. Fuzzy Control

17.1. Introduction

Much of the recent growth of interest in the field of fuzzy sets can be attributed to the success of one particular application: *fuzzy control*. While the strategy was outlined in Zadeh's paper "Outline of a new approach to the analysis of complex systems and decision processes" Zadeh (1973) the actual application was pioneered by Mamdani Mamdani and Baaklini (1975)Mamdani (1976)Mamdani (1977) and his students in the late 1970s.

Traditional control mechanisms for complex systems require advanced engineering mathematics, including the solutions of difficult differential and/or integral equations. Even worse, traditional solutions are often unsatisfactory or impossible to obtain. An excellent example of this is the recent success in Japan by Sugeno of constructing a controller for a model helicopter. This feat has not been matched by any of the traditional differential/integral controllers. The equations of motion are extremely complex and the aerodynamic interactions are difficult to express. Yet humans fly helicopters and it appears that soon fuzzy controllers will be able to as well.

Fuzzy controllers are (comparatively) easy to build. Fuzzy controllers, unlike neural network methods (addressed online in the Neural Network Chapter `http://duck.creighton.edu/Fuzzy/`), produce results that are understandable to humans. They can incorporate expert driven and data driven information. Fuzzy controllers can also be designed to adapt to changing conditions.

- Fuzzy controllers are simple to build and operate. They are based on the principles of linguistic systems and approximate reasoning discussed in Chapter (15.3). The operations used by a fuzzy controller consist of addition, subtraction, multiplication, division, maximum and minimum. Simple arithmetic operations mean that fuzzy controllers are very fast in application.

- Fuzzy controllers are easy to understand because they directly translates the actions of a human controller into simple deterministic rules; the familiar `if A then B` rules of fuzzy logic. These rules can be read back from the fuzzy controller, however it was designed, so that a human can understand what a fuzzy controller does. Thus they are based on translations of .

- Fuzzy controllers can adapt the heuristic rules we use daily to control machines such as cars and stoves. A car driving controller can be based on human rules like: "If you are going a bit too fast then apply the brakes lightly." A fuzzy controller for a factory process could use a database of past performance to build a collection of rules. A fuzzy controller for a helicopter can combine both heuristic and data driven methodologies.

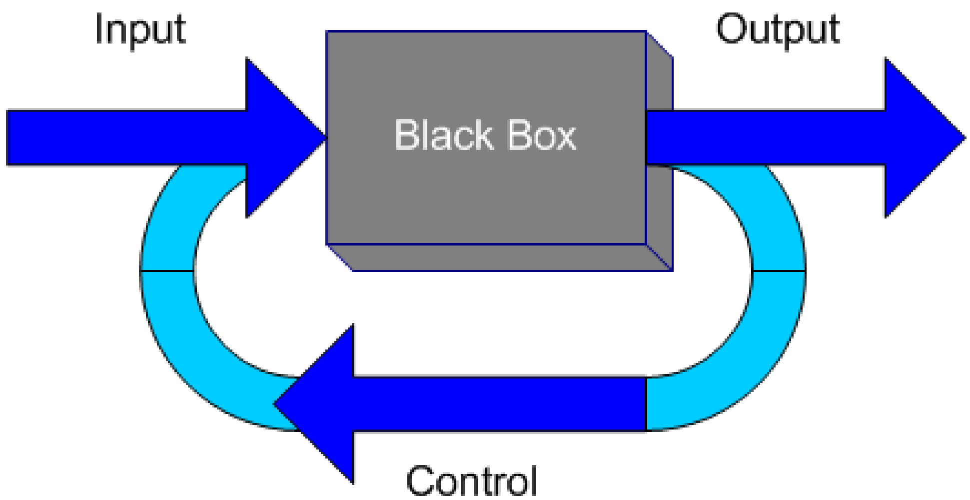


Figure 17.1.: A typical feedback control system.

- Fuzzy controllers are adaptable because the functions that represent the heuristic rules are easy to adjust using an adaptive methodology (such as a neural network or genetic algorithms).

It is important to understand that fuzzy sets provides a method for representing the inherent imprecision of the human-machine interaction, but that fuzzy sets themselves are precise mathematical functions. Probabilities present information about random processes but the probabilities are not random. Fuzzy sets represent indeterminate information but fuzzy sets are deterministic.

17.2. The black box approach

Given enough I/O data can we build a model without understanding anything about the boxes internal workings? Can we use this model to augment the input to get a desired output? This is the problem of *CONTROL*.

17.2.1. Keep on trucking

When an expert driver is trying to teach a new trainee how to back up a sixteen wheeled tractor trailer filled with large heavy machinery to a loading dock he does not give the trainee a set of complex multivariate differential equations.

$$\frac{\partial^2 y}{\partial x^2} = x \sin y - \frac{1}{2}x^2 + k\frac{\partial y}{\partial x} \tag{17.1}$$

Instead he gives the trainee a set of heuristic rules such as, “If the truck is close to the dock drive slowly” and “If the angle is large turn the steering wheel hard.” Fuzzy logic can translate these heuristic if-then rules concurrently in terms of a standard

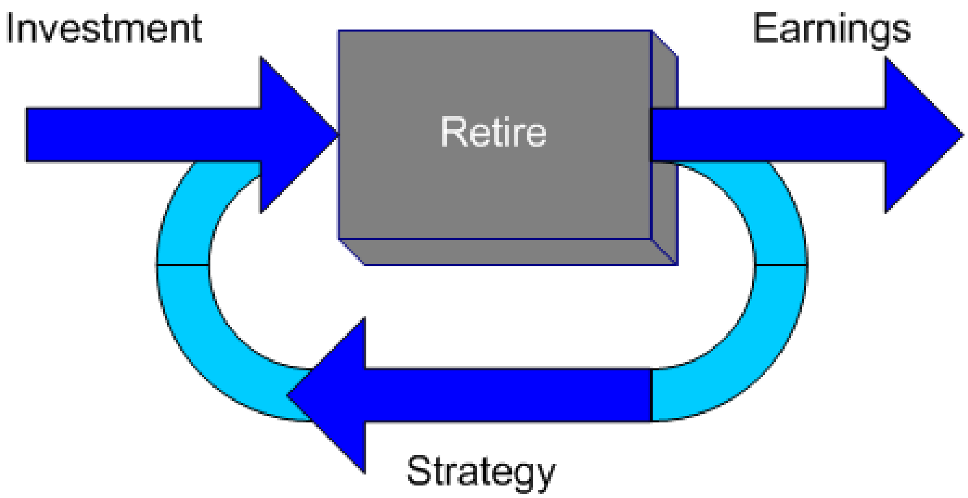


Figure 17.2.: Does this interest you?

modeling object — the fuzzy number and a standard operating system — approximate reasoning.

17.2.2. Control from Approximate reasoning

When we want to control a dam we might have a rule like "If the water level is too high open up the outlet valve a little." We might also have a reading like the water is rather high. What is needed is a numerical value for the control mechanism like "open the valve by turning the wheel an additional five degrees." There are two problems with the approximate reasoning scheme of the previous section. The input reading for a controller are crisp, such as the water is at 831 centimeters, and do not exactly match the A part of the inference rule system of approximate reasoning. Second, we do not have a connective for "implies" as we do for "and" and "or".

If we remember the previous chapters discussion about implication in traditional logical and think about the dam problem we see that the inference we want to draw is nothing at all like a deduction scheme from formal logic. For example, the table for logical implication in Table (15.1) , $\rightarrow$ says that if A is false then $A \rightarrow B$ is always true no matter what value B has. If we tried to apply this line of reasoning to the dam situation we might reason that since the water level is low it just won't matter what we do.

We are not interested in anything like this. We are not interested in a system based on truth, absolute knowledge and inflexible rules.

Another problem with traditional logic is its linearity. In the real world the data may have properties in common with many of the different heuristic rules that an expert uses to reason out the correct actions in a difficult situation. The human brain works in parallel and weighs the alternatives. When a chef looks in the refrigerator and examines then contents, the list of available ingredients may match a lot of recipes. The chef must use a complex set of rules to balance freshness of ingredients,

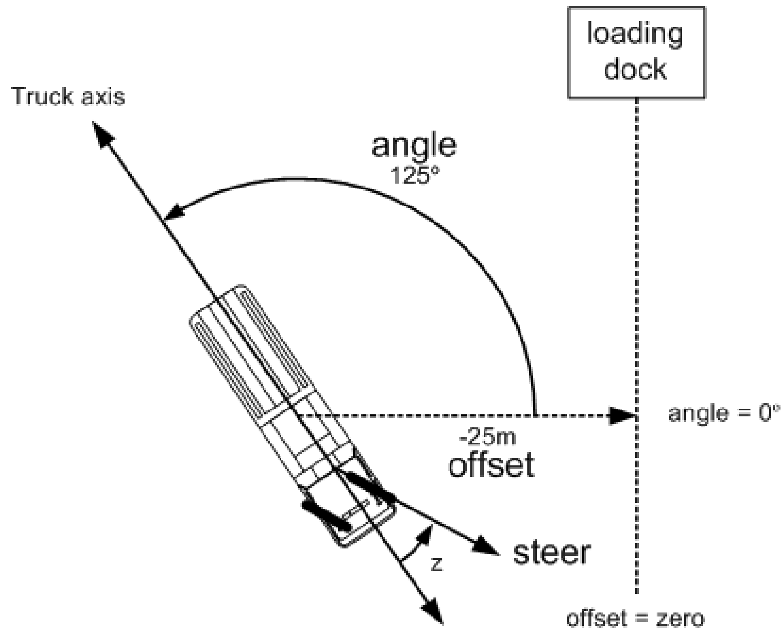


Figure 17.3.: The truck's $angle$ and $offset$ will be used as input to control the truck. The output will be the change in steering angle.

flavor balance, textural and color considerations, and the general mood of the day in deciding what to prepare. This is why most people just get take–out, such decisions are just inherently difficult. Building a system to figure out what to make for dinner would be extremely difficult. Entering the data of available items and quantities would take a lot of time, more time then it would take to make dinner, and how does one enter mood information. The computer does not understand that a Valentines day meal might be measured by a different standard, that ignores caloric overload and pampers romantic expectations. Measuring freshness, color and texture is another difficulty. There are also additional considerations like a "never steak two days in a row, except sometimes" rules to ensure variety. Then again, maybe you like to have the same meal every day and are afraid of surprises. Little Katie won't eat anything but peanut butter and jelly sandwiches or chicken nuggets. Family meal planning is a complicated system!

The simple black and white problems, like chess, are either solved or could be solved with sufficient computational power. All the grey problems, like biological systems, have proved much more difficult. Reductionism has trouble dealing with emergent properties. There is nothing in the nature of three line segments that would lead one to predict that a triangle in the plane will always contain 180 degrees. There is nothing special about a molecule like cytosine except that it forms long chains (as do a lot of other molecules) and can create a mechanism that replicate itself, hence life as we know it (cytosine is one of the four components of DNA, our genes.) Logic has trouble dealing with imprecision, with values that are not one and zero, true and false.

17.3. Fuzzy logic controllers

Figure (17.3) shows a simplified image of a truck backing up to a loading dock. Training a new driver with rules such as "If the angle of the trucks major axis is $125°$ and the truck is offset -25 meters left of the dock then turn the steering wheel so the tires are $15°$ from the track axis" will probably bring the national economy to a screeching halt; it will almost certainly fail to get the truck up against the building's loading dock. This rule is applicable for a split second, and human reflexes are probably not fast enough to implement it at the critical moment. The cognitive overload of remembering the 360 times 100 (angles versus offset) different steering angle rules necessary does not even take into account what we do if the angle is $\frac{\pi}{7}$. If we add in the velocity of the truck as a third input parameter the crisp rule based approach becomes infeasible.

Yet almost everyone can learn to drive, (though everyone who drives soon realizes that almost everyone else is doing it badly). The rules we use are linguistic and few in number. The rules for human drivers are something like:

> "If you are a little left of center and angled away to the left then turn the wheel a little to the left"
>
> "If you are a lot right of center and angled down to the left then turn the wheel a little to the right."
>
> "If you are going too fast then slow down."

Human rules do not specify the numeric ranges. These rules are learned by way of examples and negative feedback. Each human, based on there own reflexes, driving style, and vehicle type determine an optimum numeric range for the linguistic variables wheel left, truck right, etc. Think of how used to your own car you are. When you get in it to drive home after a late class you do not have to look for anything, like where the lights or windshield wipers are, something you would check out in a borrowed or rented car (I hope). There are also more than one rule to cover a situation, like driving in the rain, where standard driving behavior rules are altered by the slippery road driving rules. And somehow the brain fires all these rules *in parallel* and arrives at a proper conclusion and acts upon this conclusion.

Fuzzy control is just an approximate reasoning system with multiple, usually overlapping, rules. These rules are called the rulebase. There are some differences though. In fuzzy control the implication arrow will be modeled in a completely new. An implication like $A \to B$ is meant as a production rule. A control value (the feedback in Fig. (17.2)) is an input value which is fed to the antecedent of the implication, which is A. Based on how well the input value fits the antecedent A the controller does B, that is, the more the input is A the more we do B. This fits with intuition. A rule like "If the tomato is red then it is ripe." is used when harvesting tomatoes. The more red the color is (the input) the riper it is, and the better it will taste and the more likely we are to pick it.

You build a fuzzy logic controller (FLC) by mimicking the human controller. The speed of computer processing ensures that even a bad mimicry of the human rules produces fairly good result since the computer is correcting the actions quickly. A good controller relies on feedback, Fig. (17.1). A good fuzzy controller design produces controllers comparable to the best achieved by traditional methods. Addition-

offset			
Number	**Tag**	**English**	**Fuzzy Number**
A_1	LE	left	$Tp\,[-50,-50,-40,-15]$
A_2	LC	left center	$Tr\,[-20,-10,0]$
A_3	CE	center	$Tr\,[-5,0,5]$
A_4	RC	right center	$Tr\,[0,10,20]$
A_5	RI	right	$Tp\,[16,40,50,50]$

angle			
Number	**Tag**	**English**	**Fuzzy Number**
B_1	RB	right big	$Tr\,[-100,-45,10]$
B_2	RU	right upper	$Tr\,[-10,35,60]$
B_3	RV	right vertical	$Tr\,[45,67.5,90]$
B_4	VE	vertical	$Tr\,[80,90,100]$
B_5	LV	left vertical	$Tr\,[90,112.5,135]$
B_6	LU	left upper	$Tr\,[120,155,190]$
B_7	LB	left big	$Tr\,[170,225,280]$

steer			
Number	**Tag**	**English**	**Fuzzy Number**
C_1	NB	negative big	$Tr\,[-30,-30,-15]$
C_2	NM	negative medium	$Tr\,[-25,-15,-5]$
C_3	NS	negative small	$Tr\,[-12,-6,0]$
C_4	ZE	zero	$Tr\,[-5,0,5]$
C_5	PS	positive small	$Tr\,[0,6,12]$
C_6	PM	positive medium	$Tr\,[5,15,25]$
C_7	PB	positive big	$Tr\,[15,30,30]$

Table 17.1.: Definition of offset, angle, and steer fuzzy numbers.

ally fuzzy controllers are much easier and faster to design. Finally, in some cases traditional controllers have been too difficult to design. An example of this is an autopilot for a helicopter. The physical rules of flight for a helicopter are so complex that a physical solution is plagued with boundary value and chaos problems. A fuzzy controller is based on the experience of successful helicopter pilots.

The horizontal *offset* between the dock and center of the truck, designated x, will be one of the two inputs to our prototype fuzzy controller. The second input will be y between the perpendicular y-axis and the trucks backing up direction (you have to back up to a loading dock) the . If the dock is in the center of a 100 meter yard then the offset can range from -50 to 50 meters as the truck moves from left to right.

The granularity for the *offset* will be five and the granularity for *angle* will be seven . This gives a total of thirty-five rules for our controller.

When we say that the granularity for *offset* will be five this means that we will partition (in the fuzzy sense) the angle domain into seven overlapping fuzzy sets. *Offset*, x, is the distance in the x–dimension from the truck to the loading dock. Offset has a range from -50 to 50. The tag, English name and fuzzy numbers assigned to the

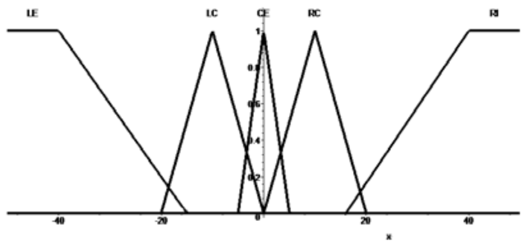


Figure 17.4.: A graph of the five fuzzy numbers LE, LC, CE, RC, and RI representing *offset* distance.

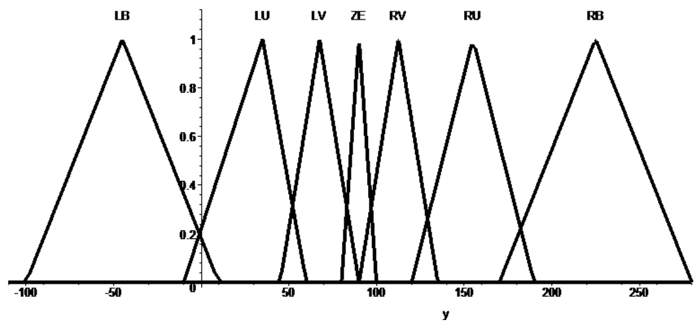


Figure 17.5.: A graph of the seven fuzzy numbers LB, LU, LV, ZE, RV, RU, and RB that partition *angle*.

five pieces of *offset* are given in Table (17.1.offset). The situation is illustrated in Fig. (17.4).

The granularity of *angle* is seven. The angle is measures as the deviation from straight up, thus the deviation angle y has a range that extends from -100 to 280 (angles are measure counterclockwise). The tag, English name and fuzzy numbers assigned to the seven pieces of *angle* are given in Table (17.1.angle). The situation is illustrated in Fig. (17.5).

The output will be the change in steering angle, designated z, to be applied by the driver to the steering wheel. The controller will use thirty–five rules to control the steering of the truck. They will approximate the heuristic rules of a human driver and, hopefully, produce a controller that mimics the actions of a human driver.

The granularity of *steer* is seven. *Steer*, z, is the change in angle of the steering wheel of the truck and has a range from -30 to 30, left to right, (its counterclockwise, steering right decreases the angle). The tag, English name and fuzzy numbers assigned to the five pieces of *steer* are given in Table (17.1.steer). The situation is illustrated in Fig. (17.6).

It will take a couple of steps to transform a heuristic rule into a fuzzy rule. Let us take the first rule from a human standpoint. "If you are way left and pointed a little

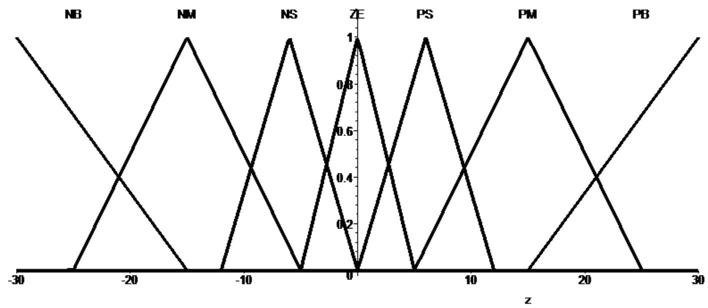


Figure 17.6.: A graph of the seven fuzzy numbers NB, NM, NS, ZE, PS, PM and PB that partition *steer*ing angle.

left (like this — ↖) then steer left."

$$\begin{array}{l}\text{“If you are pointed a little left (like this- ↖)} \\ \quad\text{and located somewhat left} \\ \quad\text{then turn the steering wheel to the left.”}\end{array} \tag{17.2}$$

The first step has already been done, we have used granulation to construct the fuzzy sets on the domain sets $X = [-50, 50]$ for *offset*, $Y = [-100, 280]$ for *angle*, and $Z = [-30, 30]$ for *steer*.

Now English sentences leave a lot to be desired as far as order and precision are concerned. To convert this English language rule to a fuzzy rule we have to convert it into a standard form. The standard form we will use is

$$A_i \wedge B_j \to C_k \tag{17.3}$$

where A_i is a fuzzy set on X, B_j is a fuzzy set on Y, and C_k is a fuzzy set on Z. This form translates as

"if *offset* is A_i and *angle* is B_j then *steer* C_k ."

The English phrase (17.2) has offset first and then angle second but our generic form is given by Eq. 17.3 which does *offset* and then *angle*, or X before Y. . It also does not have an exact match with the linguistic terms of our granulations. So let us decide that "way left" in (17.2) is associated with the *offset* fuzzy set LE or left,

A_1	LE	left	$Tp\,[-50, -50, -40, -15]$

and that "pointed a little left" is associated with the *angle* fuzzy set LV or left vertical

B_5	LV	left vertical	$Tr\,[140, 110, 90]$

.

Finally we associate with the prase "steer left" the *steer* fuzzy set PM or positive medium

C_6	PM	positive medium	$Tr\,[5, 15, 25]$

and the complete English phrase in (17.2) becomes the fuzzy rule

"If you are pointed a little left (like this- ↖)
and located somewhat left
then turn the steering wheel to the left." (17.4)

"if *offset* is left and *angle* is left vertical then *steer* positive medium." (17.5)
if *offset* is LE and *angle* is LVthen*steer* PM
$$A_1 \wedge B_5 \to C_6 \ .$$

The fuzzy controller we are going to design will need to have a decision for any possible pairing of the input *angle* and input *offset*. This means there are thirty–five rules necessary for the fuzzy controller to handle any acceptable input values in the domain $Y \times X$ (angle by offset). The best way to present them is as a table. The fuzzy controller **TRUCK** is defined in Table 17.2.

Truck Rules		B_1 RB	B_2 RU	B_3 RV	B_4 VE	B_5 LV	B_6 LU	B_7 LB
A_1	LE	NS	PS	PM	PM	PM	PB	PB
A_2	LC	NM	NS	PS	PM	PM	PB	PB
A_3	CE	NM	NM	NS	ZE	PS	PM	PM
A_4	RC	NB	NB	NM	NM	NS	PS	PM
A_5	RI	NB	NB	NM	NM	NM	NS	PS

Truck Rules		B_1 RB	B_2 RU	B_3 RV	B_4 VE	B_5 LV	B_6 LU	B_7 LB
A_1	LE	C_3	C_5	C_6	C_5	C_6	C_7	C_7
A_2	LC	C_2	C_3	C_4	C_5	C_6	C_7	C_7
A_3	CE	C_2	C_2	C_3	C_4	C_5	C_6	C_6
A_4	RC	C_1	C_1	C_2	C_3	C_4	C_5	C_6
A_5	RI	C_1	C_1	C_2	C_2	C_2	C_3	C_5

Table 17.2.: Truck fuzzy controller rulebase

Now that we have finished designing the rulebase it is time to learn how the rulebase is used to control the truck.

In Figure (17.3) the *offset* is $x = -25$ meters an the observed *angle* of deviation is $y = 125$ degrees. This is the observed data value, the input to the fuzzy control system. If we examine the the fuzzy sets A_i, see Table (17.1), we see that $x = -25$ is in the support of the fuzzy set A_1 which is LE or left. If we examine the fuzzy sets B_j, see Table (17.1.angle), we see that $y = 125$ is in the support of the fuzzy set B_5 which is LM or left medium. This means that the rule that is fired is

$$\text{if } \textit{offset} \text{ is LE and } \textit{angle} \text{ is LM then } \textit{steer} \text{ PM}$$
$$A_1 \wedge B_5 \to C_6 \ .$$

since C_6 or PM is the output of the appropriate rule from Table (17.2). The fuzzy relation for $A_1 \wedge B_5$ is modeled as a fuzzy set whose membership function at $\langle x, y\rangle$ is just the minimum of $A_1(x)$ and $B_5(y)$. Pictorially it looks like the pyramid in Figure (17.12). However there is no need to actually construct this pyramid. The height of the pyramid at $\langle x, y\rangle$ can be calculated when it is needed and only at the points where it is needed. The values $\omega_x = A_1(x)$ and $\omega_y = B_5(y)$ are called the compatibility indexes of y and x respectively. The membership grade of $\langle x, y\rangle$ in the pyramid $A_2 \wedge B_1$ is just the minimum of these two compatibility indexes. The resultant minimum value is the firing value ω.

$$\begin{aligned}
\omega &= \min[\omega_x, \omega_y] \\
&= \min\left[A_1(y), B_5(x)\right] \\
&= A_1(x) \wedge B_2(y) \\
&= \left[A_1 \wedge B_2\right](x, y)
\end{aligned}$$

Now Zadeh and Mamdani in their original design of a fuzzy logic controller used as the result of a single implication a truncated version of the fuzzy set C_6. This truncated version of C_6 has as its membership function the formula

$$\begin{aligned}
C_{1,5}(z) &= \min\left[\omega_{1,5}, C_6(z)\right] \\
&= \omega_{1,5} \wedge C_6(z)
\end{aligned}$$

where the subscript $1,5$ indicates that this compatibility index. $\omega_{1,5}$. is the result of firing fuzzy sets A_1 and B_5. This process is illustrated in Fig. (17.7).

Example 102. Since we have input from Fig. (17.3) of $\langle -25, 125\rangle$ we can calculate the actual membership grades of $A_1(-25)$ and $B_5(125)$ to produce the compatibility indexes of $\omega_x = \frac{2}{5}$ and $\omega_y = \frac{4}{9}$ and conclude that the firing value of $\langle x, y\rangle = \langle -25, 125\rangle$ is $\omega = \min\left[\frac{2}{5}, \frac{4}{9}\right] = 0.4$. Thus the resultant fuzzy set $C_{1,5}$ looks like C_6 with 60% of the top chopped off, see Fig. (17.7).

Further examination of the rules and membership grades allows us to note that $y = 125$ is also an member of the fuzzy set B_6 interpreted as LU or left upper. This means that in addition to the rule $A_1 \wedge B_5 \to C_6$ being fired the rule $A_1 \wedge B_6 \to C_7$ is fired. The result of the firing $A_1 \wedge B_6$ will be another truncated fuzzy set.

Example 103. Since we have input from Fig. (17.3) of $\langle -25, 125\rangle$ we can calculate the actual membership grades of $A_1(-25)$ and $B_6(125)$ to produce the compatibility indexes of $\omega_x = \frac{2}{5}$ and $\omega_y = \frac{1}{7}$ and conclude that the firing value of $\langle x, y\rangle = \langle -25, 125\rangle$ is $\omega = \min\left[\frac{2}{5}, \frac{1}{7}\right] = 0.14$. Thus the resultant fuzzy set $C_{1,6}$ looks like C_7 with $\frac{6}{7}$ of the top chopped off, see Fig. (17.8).

Both of the rules that are fired by $\langle x, y\rangle$, or are compatible with $\langle x, y\rangle$, are equally important. The fuzzy controller, like a human being, amalgamates all of the pertinent

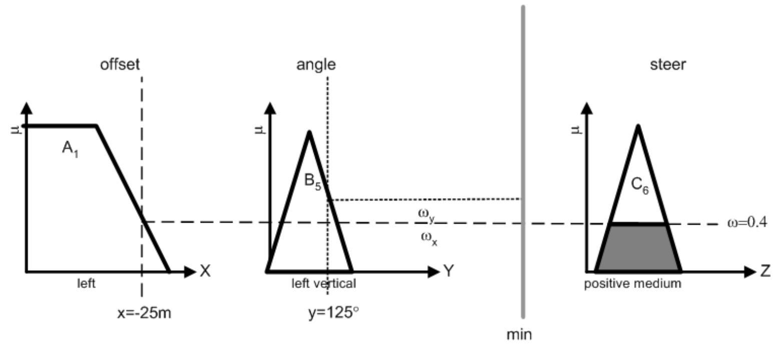


Figure 17.7.: A and B imply C in a fuzzy controller.

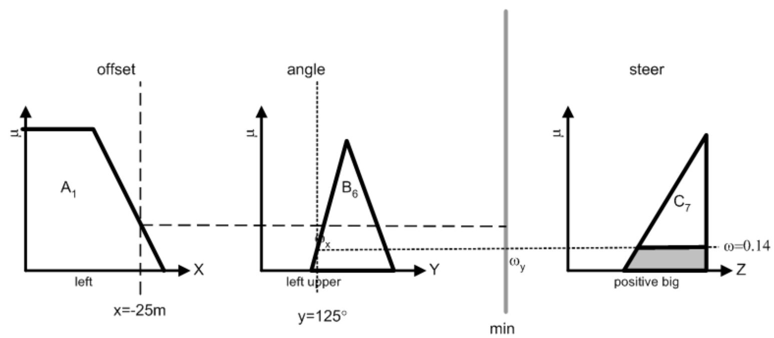


Figure 17.8.: A second if-then rule fired in parallel with the rule in Fig. (17.7).

rules. It does this by uniting (in the sense of a fuzzy set union) all of the truncated conclusions, in this case the truncated conclusions are the fuzz sets $C_{1,5}$ and $C_{1,6}$. Amalgamating the two truncated fuzzy sets with the $\max$ operator to produce the result fuzzy set C. This is the final result, the irregular fuzzy set $C(z)$ seen in figure (17.10).

The last step in the fuzzy controller is to produce a single real number to use as the control value; the number of degrees that the steering angle should be altered from its present position.

When fuzzy control was first proposed by Mamdani, he used the z value with the largest resultant membership value in the fuzzy set C as the control value. If, as in this case, the result achieves a maximum membership value over an interval $[z_1, z_2]$, then Mamdani proposed to take the average of the endpoints to produce a control value $z = \frac{z_1+z_2}{2}$.

The fuzzy set $C_{1,6}$ is a little bit taller than $C_{1,5}$ at its maximum. It achieves this maximum of 0.4 over the range $\underline{c}(0.6) = 9$ to $\overline{c}(0.6) = 21$. When we average these two values we get the control value $z_c = 15$, the actual numeric output of the fuzzy controller. The steering angle is therefore changed from its present value to 15 degrees left of vertical, which turns the wheels to the left. The process of determining a single numeric value from a fuzzy set is called defuzzification. The Mamdani controller uses MOM, mean of maxima.

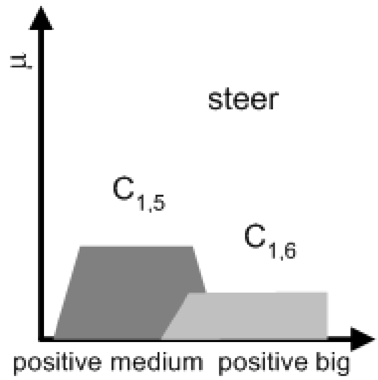


Figure 17.9.: Two rules are fired, producing two truncated triangles.

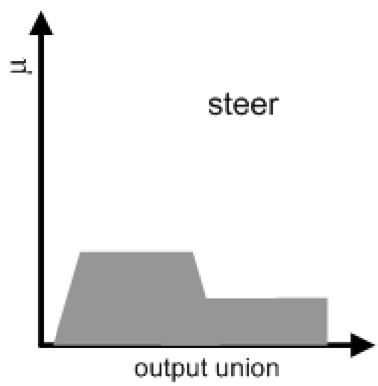


Figure 17.10.: Final fuzzy set result for steering.

We then use an electronic sensor (or other method) to get the next pair of input values $\langle x, y\rangle$ and repeat the process until the truck arrives at the dock.

The entire process of a fuzzy controller is illustrated in Fig. (17.11). The input data $\langle -25, 125\rangle$ fires two different rules in parallel. The amalgamated fuzzy set in the lower right is defuzzified to produce the control value $z_c = 15$

17.3.1. Fuzzy control equations

We finally have all the pieces for the calculation of the control fuzzy set C.

Each input data pair $\langle x_0, y_0\rangle$ is represented by an impulse fuzzy set, $A' \wedge B' = Ti[x_0] \wedge Ti[y_0]$. This impulse fuzzy set is just a single spike in the two dimensional $X \times Y$ plane at $\langle x_0, y_0\rangle$. This input membership function fires all the rules that make up the controller, but the result will be zero if the input $\langle x_0, y_0\rangle$ is not in the support of the antecedent $A \wedge B$ of the firing rule. Using the standard operations of $\max$-$\min$ composition and implication as cylindric closure we calculate the result of firing a single rule $[A_i \wedge B_j] \to C_k$. The result is $C_{i,j} = [A' \wedge B'] \circ [A_i \wedge B_j] \to C_k$, which upon a lot of algebraic manipulation produces

$$
\begin{aligned}
C_{i,j}(z) &= [A'(x) \wedge B'(y)] \circ [A_i(x) \wedge B_j(y)] \to C_k(z) && (17.6)\\
&= \sup_{x\in X, y\in Y} \min\left[A'(x) \wedge B'(y), A_i(x) \wedge B_j(y) \to C_k(z)\right] && (17.7)\\
&= \sup_{x\in X, y\in Y} \min\left[A'(x) \wedge B'(y), \min[A_i(x) \wedge B_j(y), C_k(z)]\right] && (17.8)\\
&= \sup_{x\in X, y\in Y} \min\left[A'(x) \wedge B'(y), A_i(x) \wedge B_j(y), C_k(z)\right] && (17.9)\\
&= \min\left[\sup_{x\in X}[A'(x), A_i(x)], \sup_{y\in Y}[B'(y), B(y)], C_k(z)\right] && (17.10)\\
&= \sup_{x\in X}[A'(x), A_i(x)] \wedge \sup_{y\in Y}[B'(y), B_j(y)] \wedge C_k(z) && (17.11)
\end{aligned}
$$

In this case we have that A' and B' are impulse fuzzy numbers, $A' = Ti[x_0]$ and $B' = Ti[y_0]$, so that

$$C_{i,j} = [Ti[x_o] \wedge Ti[y_0]] \circ [A_i \wedge B_j] \to C_k$$

and the final result will be

$$
\begin{aligned}
C_{i,j}(z) &= \sup_{x\in X}[Ti[x_o](x), A_i(x)] \wedge \sup_{x\in X}[Ti[y_0](y), B_j(y)] \wedge C_k(z)\\
&= A(x_0) \wedge B(y_0) \wedge C_k(z)
\end{aligned}
$$

since $Ti[x_0](x)$ is zero everywhere but $x = x_0$ and $Ti[y_0](y)$ is zero everywhere but $y = y_0$. The result $C_{i,j}$ is simply C_k truncated at the height $\min[A(x_0), B(y_0)]$, as previously stated. Since more than one rule may be fired, the ultimate result, C is the union ($\max$) of the individual outputs of each fired rule:

$$C = \bigcup_{i,j} [A' \wedge B'] \circ [A_i \wedge B_j] \to C_k$$

where i, j ranges over all the rules in the rulebase, which we assume has size $m \times n$.

17.3.2. Multiple inputs and outputs

The fuzzy controller outlined above is simplistic and may not do a very good job. In general, when backing a truck up to a dock we use more information than the angle of deviation from the perpendicular and the offset and control more quantities than just the steering wheel. For example we might include as an additional input the y–distance from the dock. If the truck is farther out from the dock then a small steering correction will have a longer time to take affect. Additionally, most drivers control a car primarily through the steering wheel and the brake/accelerator pedals. Thus we might have a system that contains rules that look like:

$$A_i \wedge B_j \wedge C_k \to F_s \wedge G_t \qquad (17.12)$$

where A_i is a fuzzy set that expresses a constraint on the dock angle, B_j is a constraint on the offset, C_k is a constraint on the velocity, and the result is the action F_s on

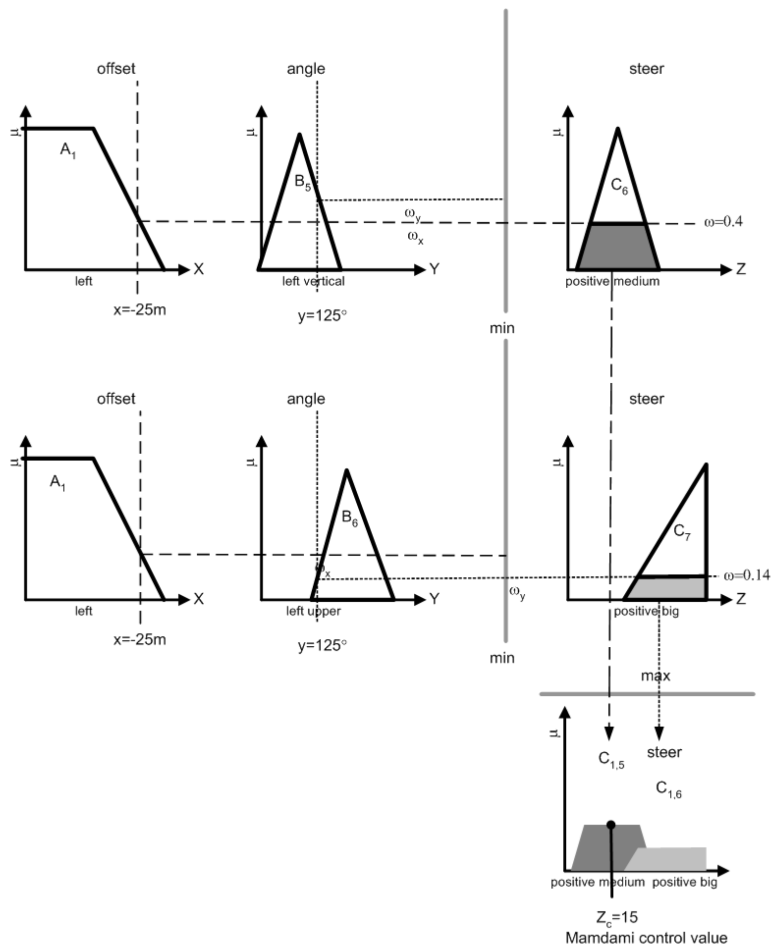


Figure 17.11.: Mamdani fuzzy logic controller.

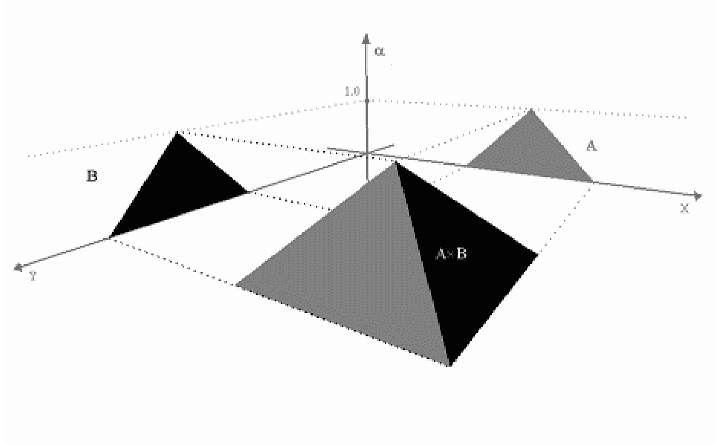


Figure 17.12.: The fuzzy set $A \wedge B$.

steering angle and G_t on acceleration (the pedals). It should be apparent that a fuzzy controller can have many antecedents as well as consequences.

The first simplification comes about when we realize that fuzzy controllers are small and fast. Instead of having one rule $A_i \wedge B_j \wedge C_k \rightarrow F_s \wedge G_t$ with multiple consequents it is usually conceptually easier to build two parallel controllers, one for the steering wheel and one for the pedals: $A_i \wedge B_j \wedge C_k \rightarrow F_s$ and $A_i \wedge B_j \wedge C_k \rightarrow G_t$.

17.3.3. Control surface

We can get a good idea of the behavior of the fuzzy logic controller we have constructed by looking at its control surface. A control surface is a 3D result of the output of the controller for each $\langle x, y \rangle$ input pair. Fig. (17.13) shows the control surface for the Truck Backer-upper described in this chapter. A fuzzy controller can be thought of as a means of approximating a function $f(x, y)$ that is the mathematically correct equation for the control value $z = f(x, y)$ given inputs x and y.

17.4. Center of gravity

When engineers, who are not so wedded to theory, examined Mamdani's controller design, there were two points that the noticed. The first was that the $\min$ operation used in truncating the then fuzzy sets was not as appealing computationally as using ω_i as a scaling factor.

Secondly, the $\max$ of membership defuzzification methodology seemed to throw away a lot of available information, as well as produce results that were not continu-

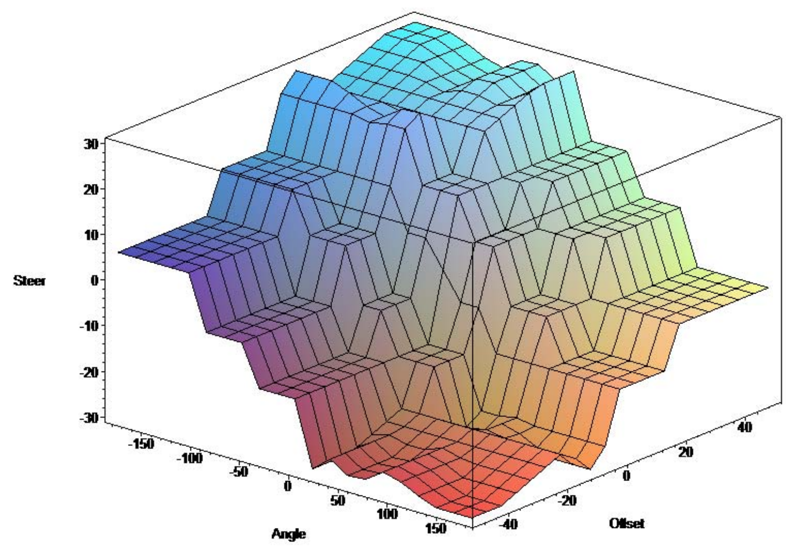


Figure 17.13.: Mamdani control surface.

ous (a small change in input angle could create a large change in the control value.) Looking at figure (17.10) an engineer would naturally choose as a typical value the *center of gravity*. The center of gravity has very nice mathematical properties as well as a simple physical explanation. The center of gravity, COG, of figure (17.10) is the balance point, the z value where we would place a fulcrum or pivot and the fuzzy set considered as a solid flat sheet of metal would then balance.

$$\mathrm{COG}(C) = \frac{\int_{-30}^{+30} z\,C(z)\,dz}{\int_{-30}^{+30} C(z)\,dz} = 16.37^\circ \tag{17.13}$$

In general, if $f : X \to \mathbb{R}$ is any function defined on a domain interval $X = [a, b]$ then the x coordinate of the center of gravity of the area between f and the x-axis is given by the formula

$$x_{\mathrm{COG}} = \frac{\int_a^b x\,f(x)\,dx}{\int_a^b f(x)\,dx}\,. \tag{17.14}$$

Here we use the variable x since this is the traditional label of the horizontal axis. However in fuzzy control, the horizontal axis of the resultant control set will be the domain set of the then fuzzy sets. In the truck example we used Z for the steering angle.

For those who do not remember calculus fondly Eq. (17.14) may appear intimidating. However, for triangular and trapezoidal fuzzy numbers we can produce simple algebraic results. For triangular fuzzy numbers $Tr\,[a, m, b]$ the COG is $\frac{a+m+b}{3}$. For

Number	**Parameters**	**COG**
Triangular	$Tr\,[a,m,b]$	$\frac{a+m+b}{3}$
Trapezoidal	$Tp\,[a,l,r,b]$	$\frac{b^2+rb+r^2-l^2-al-a^2}{3(b+r-l-a)}$

Table 17.3.: The Center of Gravity formulas for triangular and trapezoidal fuzzy numbers

trapezoidal fuzzy numbers $Tp\,[a,l,r,b]$ the COG is $\frac{b^2+rb+r^2-l^2-al-a^2}{3(b+r-l-a)}$.

Example 104. For the fuzzy number A_1 or LE for left with formula $Tp\,[-50,-50,-40,-15]$ the COG is

$$COG\,(A_1) = COG(Tp\,[-50,-50,-40,-15]) \tag{17.15}$$

$$= \frac{(-15)^2 + (-15\cdot -40) + (-40)^2 - (-50)^2 - (-50\cdot -50) - (-50)^2}{3\,(-15-40+50+50)} \tag{17.16}$$

$$= -37.592 \tag{17.17}$$

17.4.1. Why COG

The simplicity of the methodology sketched above is the basis of fuzzy control's success. The most complicated part of the procedure is the COG integrations, as given by Eq. (17.14). There are three reasons that engineers prefer COG to the original Mamdani method. The reasons are:

1. COG uses all of the output fuzzy sets of all the rules fired by a given input set.
2. The COG of two objects is the weighted sum of the COG of each object.
3. The COG of each output fuzzy set can be calculated before the fuzzy controller is run on input data.

Here is a fact from basic physics. If a planet P_1 has mass m_1 and COG x_1 and another planet P_2 has mass m_2 at COG x_2 then the center of gravity of the combined system is

$$\frac{m_1x_1 + m_2x_2}{m_1+m_2}\,. \tag{17.18}$$

This a simple formula which provides a fast method of calculating the COG of the union of the result fuzzy sets. Fast COG (FCOG) is not as precise as pure COG, since it ignores the overlap of the result fuzzy sets, but in applications of control, speed and the ability to correctly arrive at the target are more important than precision. When you drive home, you do not worry whether or not you are exactly in the middle of your driving lane, only that you are somewhere near the middle of your driving lane.

If a fuzzy controller is designed using the FCOG defuzzification method then the picture in Fig. (17.11) is not accurate. Instead of truncating the result fuzzy sets at the compatibility index ω, COG scales the fuzzy sets by the value ω. In one sense the difference between the resulting fuzzy set is the difference between using the t-norms $\min$ and *algebraic product*.

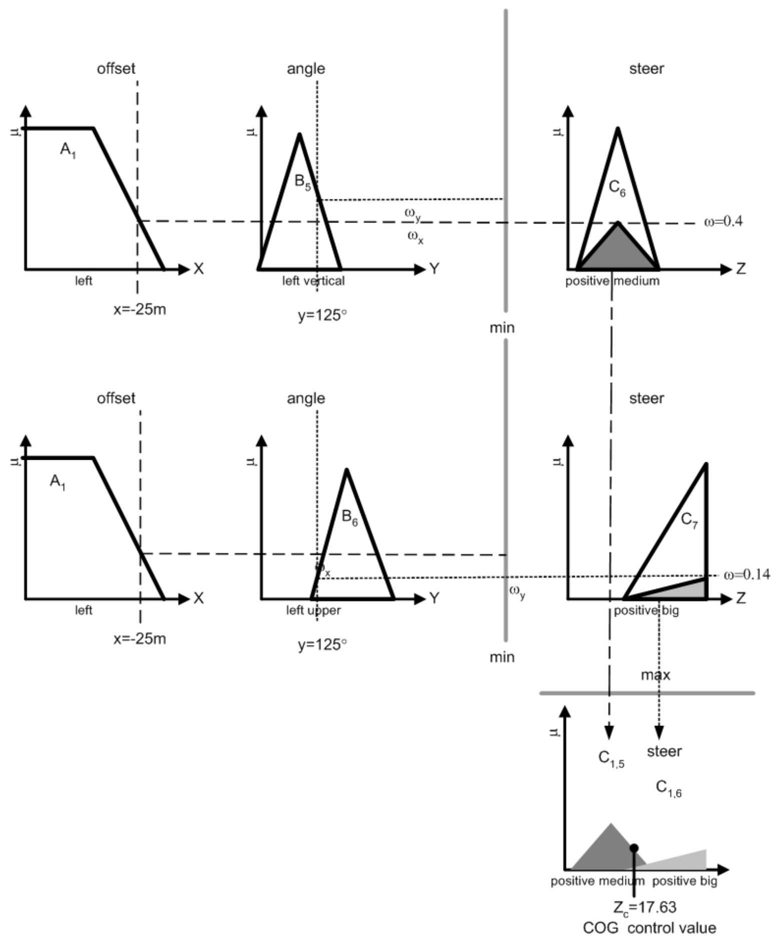


Figure 17.14.: COG fuzzy logic controller.

The situation is illustrated in Fig. (17.14).

What this means for the design and implementation of a fuzzy controllers is, calculate the COG of all the output fuzzy sets C_k before we run the fuzzy controller and store them in a table. Suppose fuzzy number C_1 has COG c_1, fuzzy number C_2 has COG c_2, etc. Further suppose that is there are rules $A_{i_1} \wedge B_{j_1} \to C_1$ and $A_{i_2} \wedge B_{j_2} \to C_2$ that are both fired by the input vector $\langle x_0, y_0 \rangle$. If the compatibility index of firing $A_{i_1} \wedge B_{j_1}$ with input $\langle x_0, y_0 \rangle$ is ω_1 and the the compatibility index of firing $A_{i_2} \wedge B_{j_2}$ with input $\langle x_0, y_0 \rangle$ is ω_2, then the resultant control value is

$$z_c = \frac{\omega_1 c_1 + \omega_2 c_2}{\omega_1 + \omega_2} \tag{17.19}$$

if these are the only two rules that are fired by the input pair $\langle x_0, y_0 \rangle$. This means that defuzzification takes two multiplications, two additions and one division after the values c_1 and c_2 are retrieved from a lookup table.

For the fuzzy controller illustrated in this chapter at most four fuzzy rules can have positive compatibility index. This is because each granulation has at most two fuzzy sets that overlap at any domain value. Since there are two inputs and each input can have two compatible fuzzy sets there are $2 \times 2 = 4$ fired rules.

Example 105. The COGs of C_6 and C_7 are 15 and 25 respectively. Assume the input to the controller is $\langle -25, 125 \rangle$. We have already calculates that $\omega_1 = 0.56$ and $\omega_2 = 0.6$ for these input values. The fast COG defuzzification control value is

$$\begin{aligned} z_c &= \frac{\omega_1 c_1 + \omega_2 c_2}{\omega_1 + \omega_2} \\ &= \frac{0.56 \cdot 25 + 0.6 \cdot 15}{0.56 + 0.6} \\ &= 17.63 \ . \end{aligned}$$

When we use fast COG the control surface for the Truck Backer-upper is illustrated in Fig (17.15).

17.5. Notes

While Zadeh created the linguistic systems of approximate reasoning and pretty much spelled out how a fuzzy controller should work in papers such as Zadeh (1973)Zadeh (1972b) Mamdani Mamdani and Assilian (1975) gets the credit of the first paper that describes a theoretical fuzzy controller.

Fuzzy set theory was suffering from one of the periodic AI implosions, and researchers in the west did not see the point of fuzzy controllers. The Japanese seized upon the technology, which made a major impression when it was implemented in the Sendai subway system. Subsequently the Japanese incorporated FLCs into their cars and electronics, and made a huge profit.

There are innumerable books on fuzzy control, but none better at the introductory level, I hope, than the one you are reading. At an advanced level Nguyen et al. (1995) has some good papers and your best bet is to browse the journals *Fuzzy Sets and Systems* and *IEEE Transactions on Fuzzy Systems*.

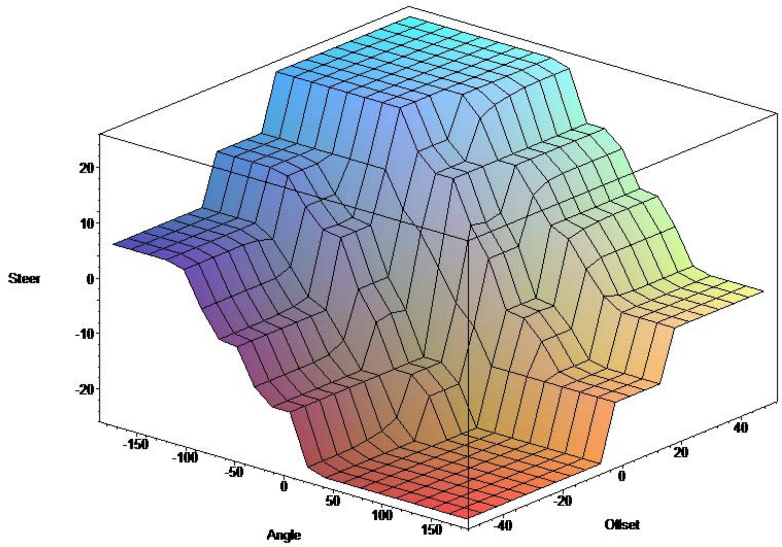


Figure 17.15.: COG control surface.

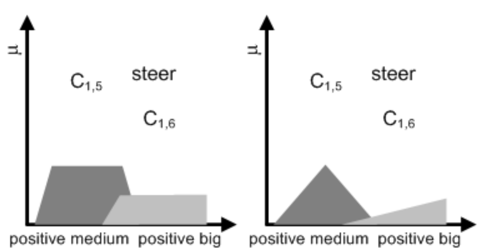


Figure 17.16.: Resultant fuzzy sets using: left – cutting and right – scaling.

17.6. Homework

Given the fuzzy controller described in Sec. (17.2).

1. What is the COG of the seven fuzzy numbers in Fig. (17.6).
2. What is the COG of the five fuzzy numbers in Fig. (17.4).
3. What is the COG of the seven fuzzy numbers in Fig. (17.5).
4. What is the compatibility index with an input of $\langle x, y\rangle = \langle 10, -5°\rangle$ of each of the rules this input triggers in Table 17.2.
5. Graph the resultant fuzzy sets of the controller upon input of $\langle x, y\rangle = \langle 10, -5°\rangle$ if truncation is used.
6. Graph the resultant fuzzy sets of the controller upon input of $\langle x, y\rangle = \langle 10, -5°\rangle$ if scaling is used.
7. Defuzzify the result of (5) using Mamdani's original method.
8. Suppose that the truck is moving one meter per second. What is the approximate location of the truck one second after the fuzzy controller alters the steering angle if we use the defuzzification value of the previous question?
9. Defuzzify the result of (6) using the COG method.
10. Suppose that the truck is moving one meter per second. What is the approximate location of the truck one second after the fuzzy controller alters the steering angle if we use the defuzzification value of the previous question?
11. What is the compatibility index with an input of $\langle x, y\rangle = \langle -10, 100°\rangle$ of each of the rules this input triggers.
12. Graph the resultant fuzzy sets of the controller upon input of $\langle x, y\rangle = \langle -10, 100°\rangle$ if truncation is used.
13. Graph the resultant fuzzy sets of the controller upon input of $\langle x, y\rangle = \langle -10, 100°\rangle$ if scaling is used.
14. Defuzzify the result of (5) using Mamdani's original method.
15. Suppose that the truck is moving one meter per second. What is the approximate location of the truck one second after the fuzzy controller alters the steering angle if we use the defuzzification value of the previous question?
16. Defuzzify the result of (6) using the COG method.
17. Suppose that the truck is moving one meter per second. What is the approximate location of the truck one second after the fuzzy controller alters the steering angle if we use the defuzzification value of the previous question?
18. Look at the resultant fuzzy sets in Fig. 17.11, reproduced in the left of Fig. (17.16). What values besides the MOM could be used for defuzzification.
19. Look at the resultant fuzzy sets in Fig. 17.14, reproduced in the left of Fig. (17.16).. What values besides the COG could be used for defuzzification.

18. Fuzzy Logic Controllers

18.1. Variations on a theme by Mamdani

Designing fuzzy logic controllers (FLC) is not difficult. Building optimal fuzzy logic controllers, with sufficient speed or accuracy, is a bit more difficult. Fuzzy controllers have proven to be so effective that they are in the great majority of consumer electronic devices. For example, the dryer in my basement has a moisture sensor, and lowers the temperature of the hot air as the clothes approach the optimal dryness, this prevents shrinkage. This process is almost certainly controlled by a FLC, but this is never stated because in America, the words fuzzy logic, like fuzzy math, are considered pejorative. However, fuzzy logic brakes make driving a car a more pleasant experience. FLCs help land the space shuttle and aim the gun of an Abrams tank.

The success of FLCs has caused a lot of researchers to examine, and tweak the methods used in their construction and execution. There are four major components of a FLC. FLC construction consists of *granulation* and *rulebase* formation. FLC execution consists of an *inference* engine and *defuzzification*.

18.2. The six components

Granulation Determining the domain sets, the shapes, and the quantity, of input and output fuzzy sets. This step results in fuzzification.

Fuzzification The construction of fuzzy numbers associated with linguistic tags to represent the potential data input and control output values.

Rulebase Constructing firing rules with fuzzy numbers and logical connectives ($\tilde{}, \vee, \wedge, \rightarrow$) to represent the knowledge base.

Operators Determining the appropriate fuzzy operators to model each logical connective.

Inference Transform the input data into fuzzy numbers. Compose the observed input data fuzzy number with the set of implication in the rulebase to derive a conclusion fuzzy set. This is also called firing the rules.

Defuzzification Determining a single control value from conclusion fuzzy set.

There are thousands of products out there with fuzzy controllers. These include the autofocus cameras, motioncontrol camcorders, washing machines that sense moisture, elevators that stop smoothly, and automobile transmission and braking systems. Fuzzy controllers are a part of everyday life.

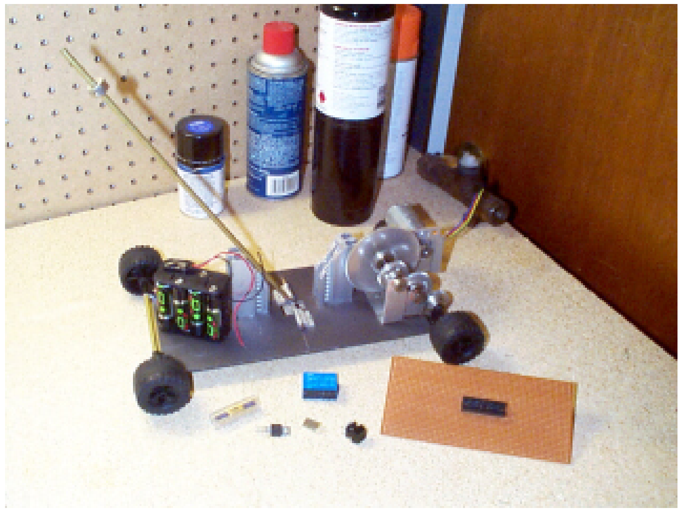


Figure 18.1.: A cart with an inverted pendulum.

Example 106. This example gives a hint of how to do create an inverted pendulum controller the traditional way. If you don't make it through this physics problem that is okay. The point of this example is the complexity of the setup of a controller for a very uninteresting problem for the fuzzy engineer. The fuzzy engineer can discard the physics and deal with it in a human way. However, for very interesting problem like flying a space shuttle or a helicopter the physical description becomes important..

The cart with an inverted pendulum, shown in Fig. (18.1) is "bumped" with an impulse force, F as illustrated in Fig. (18.2).

For this example, let's assume that the physical constants and variables are defined as follows:

M mass of the cart $M = 0.5$ (kg),

m mass of the pendulum $m = 0.2$ (kg),
b friction of the cart $b = 0.1$ ($\frac{\text{N}}{\text{m/sec}}$),
I inertia of the pendulum $I = 0.006$ (kg·m^2),
l length to the pendulum's center of mass $l = 0.3$ (m),
F impulse force applied to cart,
θ angle of the pendulum, and
x x-axis location.

The cart has a small motor that can be used to move the carts left or right. For this problem, we are only interested in the control of the pendulum's position. Therefore, none of the design criteria deal with the cart's position. The system starts at equilibrium, and experiences an impulse force F of 1 N. The pendulum should return to its equilibrium upright position as a result of the motors efforts.

If we were physics, engineering or applied mathematicians then we could perform the following analysis.

Let N be the horizontal force of the pendulum. The sum of all the forces on the cart

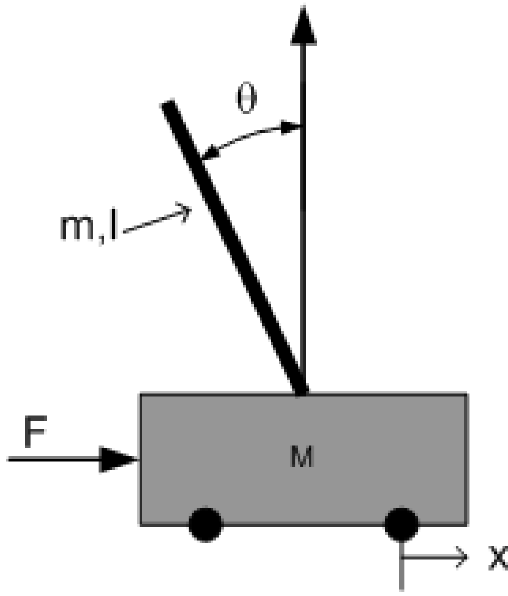


Figure 18.2.: Parameters used in the PID controller of an inverted pendulum,

in the horizontal dimension is given by:

$$M\ddot{x} + b\dot{x} + N = F$$

where $\dot{x}$ and $\ddot{x}$ represent the velocity and acceleration (first and second derivatives of the horizontal distance x).

Summing the forces in the pendulum of the cart in the horizontal direction, you get the following equation of motion:

$$N = m\ddot{x} + ml\ddot{\theta}\cos\theta + ml\ddot{\theta}^2\sin\theta$$

where $\ddot{\theta}$ represent the angular acceleration (second derivatives of the angle θ).

If you substitute this equation into the first equation, you get the first equation of motion for this system:

$$(M+m)\,\ddot{x} + b\dot{x} + m\ddot{x} + ml\ddot{\theta}\cos\theta + ml\ddot{\theta}^2\sin\theta = F$$

To get the second equation of motion, sum the forces perpendicular to the pendulum and calculate the following equation:

$$\left(I + ml^2\right)\ddot{\theta} + mgl\sin\theta = -ml\ddot{x}\cos\theta$$

These two equations, along with the initial conditions form a system of differential equations. Now that the problem has been set up it so that it can be solved or simulated on a computer to create a PID – Proportional-Integral-Derivative – controller for the servomotor.

Fortunately we do not need to remember any of this physics. We instead will use two very simple human rules, also called heuristics. Rule 1: if the pendulum is going left push right and vice versa. Rule 2: if the pendulum is speeding up push harder.

variable			
Number	**Tag**	**English**	**Fuzzy Number**
1	NB	negative big	$Trapeziodal$
2	NM	negative medium	$Triangular$
3	NS	negative small	$Triangular$
4	ZE	zero	$Triangular$
5	PS	positive small	$Triangular$
6	PM	positive medium	$Triangular$
7	PB	positive big	$Trapeziodal$

Table 18.1.: Generic captions for seven control fuzzy numbers.

18.3. Granulation

18.3.1. Domain definition

Though the first step, domain specification, is not in any sense fuzzy, it is essential to problem solving. In the truck example of the previous chapter, the angle input ranges over the full $360°$ that the truck might be oriented. However, the interval Y does not start at $0°$ and end at $360°$. That is because the target for the trucks angle is straight up, $90°$ and it turns out that it is desirable to have the target in the middle of the input range, though it is not necessary. The FLC designer needs to consider what happens if the angle is straight down, which could be $-90°$ or $270°$ depending on your viewpoint. The rules need to cover this situation and it turns out that the best practical solution is to set $Y = [-100°, 280]$ so that the controller can view it either way and have a rule that covers each situation. The idea is, if the angle is $269°$ and increases a little to $273°$ the controller does not convert this to $-87°$ and use a different strategy (rule in the rulebase) that causes the truck to jerk back and forth.

Example (106) shows the type of mathematics and physics that a typical engineer needs to design a traditional PID controller. Fig. (18.3) shows a mechanical system by ADWIN corporation for designing and testing fuzzy controllers. The ADWIN system dispenses with the cart and uses a servomotor that directly applies torque to the pendulum. The controller, after firing and defuzzyfying the rules, eventually applies a voltage to the servomotor (which can also be seen on the cart in Fig. (18.1)) that attempts to get the pendulum to a vertical position.

Let us consider the design of a controller for a fixed inverted pendulum. One of the wonderful things about fuzzy controller design is happy ignorance. We are going to ignore almost everything about physics and mathematics and focus on two very generic quantities, *error*, e, and rate of error change, Δe. The goal of the inverted pendulum FLC is to drive e and Δe to zero.

The inverted pendulum FLC will have two inputs and one output. The output is the correction factor or the feedback. Straight up, in radians, is $\frac{\pi}{2}$ and this is the target *location*, we want the error e to be zero when the pendulum is straight up at location $\frac{\pi}{2}$. The target of Δe, the *rate of change of the error*, will also be zero when the error is constant. In physical terms e is the *angular deviation* and Δe is the *angular velocity*. When both are zero the pendulum is straight up and not moving.

Assumes that the controller runs continuously and can produce input samples at

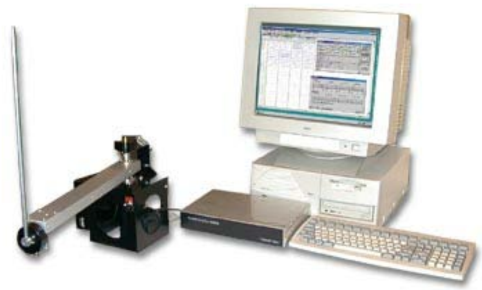


Figure 18.3.: Adwin corporations commercial fuzzy controller system.

the rate of once per second. At each clock tick the mechanical sensor needs to read the current pendulum angle θ. The value of the error variable e is the deviation from vertical. The error is then the distance form θ to $\frac{\pi}{2}$

$$e = \frac{\pi}{2} - \theta \; . \tag{18.1}$$

The difference between two readings divided by the time period is the angular velocity $\Delta\theta$. Since we wisely set the time period to one second we do not need to bother to divide by one, though technically we should do this to get the physical unit of radians per second for $\Delta\theta = \Delta e$.

We do not need a sensor for the rate of change of error Δe. Initially set Δe to zero. As soon as we get the second reading we use very simple calculation, Δe is equal to the current value of e minus the previous value of e :

$$\begin{aligned}\Delta e &= e - e_{last}\\ &= \theta - \theta_{last} \; .\end{aligned}$$

Example 107. Suppose that you buy a ten pound turkey. You estimate that this turkey will cook in five hours at $450°$. Your goal is to have the turkey at the correct internal temperature when the guests are ready for dinner. There is no need to alter the temperature for the first four hours. It is only in the last hour that we tweak the oven settings based on the current internal temperature of the turkey (they good ones come with an embedded thermometer) and how fast it seems to be cooking. The current internal temperature minus the goal internal temperature is e and the difference between the last two values of e is Δe.

Similar to a human, a FLC has course rules for great deviations and fine rules for small deviations.

This thinking influences the ranges of the variables e and Δe and the granulation. If either the location error e or velocity error is Δe is a large value then all we can do to balance the pole is crank the servomotor to full positive or negative voltage and hope the servomotor is strong enough to overcome the pendulums momentum. It is close to the target values of location error $e = 0$ and velocity error $\Delta e = 0$ where we need to use fine control to stabilize the inverted pendulum at the vertical position.

The domain of the error e will be numerically $E = \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ but remember that the clockwise orientation of angles make this right to left. The domain range for Δe is the same numerically $\Delta E = \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ but this is physically the angular velocity, also right to left.

The output of the controller is the DC current to the servomotor. According to the package the range is $\pm 10v$. So the control variable z will have domain $Z = [-10, 10]$.

18.3.2. Shape determination

Chapter (7) gave many different shape prototypes for fuzzy numbers and s-shaped fuzzy sets. Most FLC applications use four basic types of fuzzy numbers, impulse, triangular, trapezoidal, and bell. Bell shaped fuzzy number are primarily used when calculus will be involved in the design of the fuzzy controller. Bell shaped fuzzy number are differentiable and integrable, whereas triangular fuzzy numbers are not differentiable since they have corners.

The simplest and fastest method of designing a fuzzy controller is to just use equally spaced triangular fuzzy numbers. like in Fig. (18.4). Better results usually come from converting the fuzzy numbers at the extreme ends of the interval to trapezoidal fuzzy numbers, as illustrated in Fig. (18.5). The best results come from the situation illustrated in Fig. (18.6) where the fuzzy numbers at the target (the center in this case) are narrower, allowing for fine control near the target.

In the design of our fuzzy logic controller for the inverted pendulum, we will use the triangular numbers in the center of the domain intervals, trapezoidal numbers at the edges of the domain interval, and impulse fuzzy numbers for the input data.

18.3.3. Number

In FLC design the rule of thumb is that an odd number of fuzzy sets will be spread across each of the variables domain intervals. In general, the larger the domain, the more fuzzy sets that are needed. However, there are demonstrations of a FLC for the inverted pendulum that divides each of the domain sets, E, ΔE, and Z, into three pieces. This FLC works fine, it is just not very efficient.

For some reason, a granulation of five, seven, or nine pieces seems serviceable in most applications. To make life easier we will use seven fuzzy sets for each of the three domain intervals. Since, by clever design, all of the intervals are symmetric about zero we will use identical generic linguistic tags for each of the three variables e, Δe, and z. They are given in Table (18.1).

18.3.4. Fuzzification

Various empirical experiments in FLC design have shown that the granules of each domain interval should have a $10\% - 20\%$ overlap. This is a heuristic rule and can be used to produce a fast first approximation of an optimal controller. In addition, we have already decided to use seven fuzzy sets, with generic names. The central five

error			
Number	**Tag**	**English**	**Fuzzy Number**
E_1	NB	negative big	$Tp\left[-\frac{\pi}{2}, -\frac{\pi}{2}, -\frac{3\pi}{8}, -\frac{9\pi}{32}\right]$
E_2	NM	negative medium	$Tr\left[-\frac{5\pi}{16}, -\frac{\pi}{4}, -\frac{3\pi}{16}\right]$
E_3	NS	negative small	$Tr\left[-\frac{7\pi}{32}, -\frac{\pi}{8}, -\frac{3\pi}{32}\right]$
E_4	ZE	zero	$Tr\left[-\frac{\pi}{8}, 0, \frac{\pi}{8}\right]$
E_5	PS	positive small	$Tr\left[\frac{3\pi}{32}, \frac{\pi}{8}, \frac{7\pi}{32}\right]$
E_6	PM	positive medium	$Tr\left[\frac{3\pi}{16}, \frac{\pi}{4}, \frac{5\pi}{16}\right]$
E_7	PB	positive big	$Tp\left[\frac{9\pi}{32}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{\pi}{2}\right]$

change in error			
Number	**Tag**	**English**	**Fuzzy Number**
ΔE_1	NB	negative big	$Tp\left[-\frac{\pi}{2}, -\frac{\pi}{2}, -\frac{3\pi}{8}, -\frac{9\pi}{32}\right]$
ΔE_2	NM	negative medium	$Tr\left[-\frac{5\pi}{16}, -\frac{\pi}{4}, -\frac{3\pi}{16}\right]$
ΔE_3	NS	negative small	$Tr\left[-\frac{7\pi}{32}, -\frac{\pi}{8}, -\frac{3\pi}{32}\right]$
ΔE_4	ZE	zero	$Tr\left[-\frac{\pi}{8}, 0, \frac{\pi}{8}\right]$
ΔE_5	PS	positive small	$Tr\left[\frac{3\pi}{32}, \frac{\pi}{8}, \frac{7\pi}{32}\right]$
ΔE_6	PM	positive medium	$Tr\left[\frac{3\pi}{16}, \frac{\pi}{4}, \frac{5\pi}{16}\right]$
ΔE_7	PB	positive big	$Tp\left[\frac{9\pi}{32}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{\pi}{2}\right]$

control			
Number	**Tag**	**English**	**Fuzzy Number**
C_1	NB	negative big	$Tp\,[-10, -10, -8, -8]$
C_2	NM	negative medium	$Tr\,[-8, -4, -6]$
C_3	NS	negative small	$Tr\,[-5, -3, -1]$
C_4	ZE	zero	$Tr\,[-2, 0, 2]$
C_5	PS	positive small	$Tr\,[1, 3, 5]$
C_6	PM	positive medium	$Tr\,[4, 6, 8]$
C_7	PB	positive big	$Tp\,[7, 8, 10.10]$

Table 18.2.: Definition of error, delta-error, and control fuzzy numbers.

fuzzy numbers will be triangular and the extreme fuzzy numbers will be trapezoidal. Using this strategy and a little hand arithmetic produces the granulations for E, ΔE, and Z given in Table (18.2).

A typical approach to granulation will produce the following:

18.4. Rulebase

Once the components, the fuzzy numbers are available, it is time to construct a rule base. Alternatively, it can be said that it is time to construct the set of logical statements that constitute our expert systems.

18.4.1. Rulebase construction methods

There are essentially four different methodologies for the construction of a FLC rule base.

18.4.1.1. Expert experience and control engineering knowledge

Fuzzy control rules have the form of fuzzy conditional statements that relate the state oft he system variables in the antecedent — the if part — and process control variables in the consequents — the then part .In our daily life, most of the information on which our decisions are based comes to us in linguistic form rather than in numerical form. You can drive your car home with a broken speedometer without trouble. From this perspective, fuzzy logic rules provide a natural framework for the characterization of human behavior and for capturing the reasoning employed in decisions analysis.

Many experts have expressed satisfaction with the way that a fuzzy control rule captures their domain knowledge.

18.4.1.2. Operator's control actions

The first commercial FLC was built for a cement kiln. This was controlled by a human whose sole technological device was a piece of smoked glass used to peer at the kiln during its cycle. In many such industrial man-machine control system, the input-output relations are difficult to express in the form of differential equations or other mathematical formalism used in classical control theory for modeling and simulation. The human controller cannot express his actions with sufficient precision and no scientist has performed a detailed physical analysis of the system's components.

Yet a skilled human operators can control a cement kiln or other industrial systems successfully without having any quantitative models in their mind. A human operator employs, consciously or subconsciously, a set of rules to control the process. These internal rules can be converted to fuzzy if-then statements using linguistic terms by direct observation of the human controller's actions in terms of the input-output operating data. It should be noted that this can often take a lot of time to accomplish, and is very difficult if the operator is not cooperating in the exposition.

18.4.1.3. Fuzzy model of a process

In the linguistic approach, the linguistic description of the dynamic characteristics of a controlled process may be viewed as a fuzzy model of the process. This is what we have developed in this chapter, a fuzzy model of a pendulum system.

Let us suppose that the pendulum is a little to the left and moving left. Then we want to push it to the right. If it is moving fast to the left we push hard to the right, if it is moving slowly to the left we push softly to the right.

Based on the fuzzy model, we can generate a set of fuzzy control rules for attaining optimal performance of the dynamic system. The set of fuzzy control rules forms the rule base of an FLC. Although this approach is somewhat more complicated, since we must have some understanding of the system, it yields better performance and reliability, and provides a FLC directly.

18.4.1.4. Learning

Many fuzzy logic controllers have been built to emulate human decision-making behavior, but few are focused on human learning, namely, the ability to create fuzzy control rules and to modify them based on experience.

The major attempts to build systems that learn fuzzy rules are based on Neural Networks and Genetic Algorithms (Available online at `http://duck.creighton.edu/Fuzzy`)). Both of these methods are important enough to have their own chapter.

Example 108. With the linguistic terms of Table (18.2) we build a fuzzy model for the pendulum.

Let us suppose that the pendulum is a little to the right and moving left. If it is moving fast to the left we push medium to the right, if it is moving slowly to the left we push softly to the right. An analysis of the system might produce the rules presented in Table (18.3).

Pendulum Rules			**change in error**						
			ΔE_1	ΔE_2	ΔE_3	ΔE_4	ΔE_5	ΔE_6	ΔE_7
			NB	NM	NS	ZE	PS	PM	PB
	E_1	NB	PB	PB	PB	PB	PM	PS	ZE
	E_2	NM	PB	PM	PM	PM	PS	ZE	NS
	E_3	NS	PB	PM	PS	PS	ZE	NS	NM
error	E_4	ZE	PB	PM	PS	ZE	NS	NM	NB
	E_5	PS	PM	PS	ZE	NS	NS	NM	NB
	E_6	PM	PS	ZE	NS	NM	NM	NM	NB
	E_7	PB	ZE	NS	NM	NB	NB	NB	NB

Table 18.3.: Linguistic version of an Error-Delta-Error fuzzy controller rulebase

The table summarizes forty-nine rules. A typical example of a rule is

$$\text{if } e \text{ is NM and } \Delta e \text{ is PS then } control \text{ is PS} \tag{18.2}$$

or

$$E_2 \wedge \Delta E_5 \rightarrow Z_5 \tag{18.3}$$

18.4.2. Larsen and Sugeno controllers

There is four important successful departure from the standard design of fuzzy controllers that will be discussed in the following sections. Larsen controllers use the king of fuzzy logic rules that are used in Mamadani controllers and were presented in the previous chapter. Rules like $A(x) \wedge B(y) \rightarrow C(z)$ where A, B, and C are fuzzy sets. Since a fuzzy set is essentially a function, in a broad sense this rules can be written $A(x) \wedge B(y) \rightarrow C[\omega](z)$ where ω is the compatibility of the antecedent $A(x) \wedge B(y)$ with the input vector $\langle x_0, y_0 \rangle$. The result of this rule is a function C_ω that maps the output domain z to the unit intervals, $C_\omega : Z \rightarrow [0, 1]$. In both the Mamdani and Larsen controller $C_\omega = C \wedge \omega$. In a Mamdani controller and is modeled with $\min$ and in the Larsen controller and is modeled with multiplication.

Tsukamoto and Sugeno style controllers replace the consequent fuzzy sets C with functions that are not necessarily fuzzy numbers. Tsukamoto use rules of the type $A(x) \wedge B(y) \to C_k$ where C_k is an sigmoid function defined on the output space Z. Sugeno proposed rules of the type $A(x) \wedge B(y) \to f(x, y)$ where f is an arbitrary function. In the Sugeno case f is a function from $X \times Y$ to the real numbers, $f : X \times Y \to \mathbb{R}$.

Pendulum Rules		**Y**						
		B_1	B_2	B_3	B_4	B_5	B_6	B_7
	A_1	C_7	C_7	C_7	C_7	C_6	C_5	C_4
	A_2	C_7	C_7	C_6	C_6	C_5	C_4	C_3
	A_3	C_7	C_6	C_5	C_5	C_4	C_3	C_2
X	A_4	C_7	C_6	C_5	C_4	C_3	C_2	C_1
	A_5	C_6	C_5	C_4	C_3	C_3	C_2	C_1
	A_6	C_5	C_4	C_3	C_2	C_2	C_1	C_1
	A_7	C_4	C_3	C_2	C_1	C_1	C_1	C_1

Table 18.4.: Abstract version of an Error-Delta-Error fuzzy controller rulebase

18.5. Representation

The third step in the design of a fuzzy controller is deciding on the fuzzy operators that will be used to model the logical operators. The logical connectives are $\tilde{}$, $\vee$, $\wedge$, and $\to$. The standard interpretation is that negation, $\tilde{}$, is complement and uses the $1-\mu(x)$ operator, that logical or, $\vee$, is union and uses the $\max$ operator and that logical and,$\wedge$, is intersection and uses the $\min$ operator. Remember that union or $\max$ is used to amalgamate the outputs of each of the individual rules.

The success of fuzzy logic controllers has spurred researchers to examine a lot of alternatives to the standard model for the logical connectives. This is usually done by using a negation operator c for complements, a t-conorm for union and a t-norm for intersection. Some researchers have tried fairly arbitrary operators for $\vee$ and $\wedge$ including t-norms, t-conorms, and aggregation operators.

If we examine the control equations

$$C_{i,j} = [A' \wedge B'] \circ [A_i \wedge B_j] \to C_k \tag{18.4}$$

(see Eq. (17.6)) we see that in the original Mamdani FLC, as well as most production controllers $\to$ is modeled with the t-norm $\min$ which is not properly an implication operator (see Sec. (15.1)).

Mizumoto (1988) gives a thorough examination of using an arbitrary implication operator i to model logical implication.

18.6. Inference

The FLC has at this stage been designed. It now needs to be applied, to be used, so the next step is the firing of the rule base, or *inference*. This step produces the

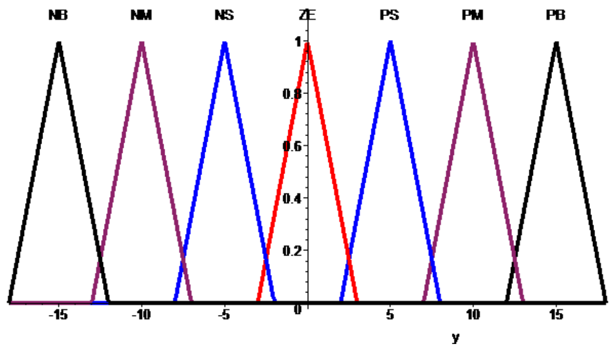


Figure 18.4.: Granulation with triangular fuzzy numbers.

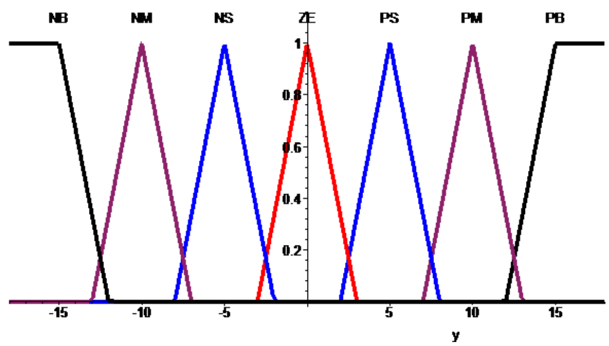


Figure 18.5.: Granulation with triangular and trapezoidal fuzzy numbers.

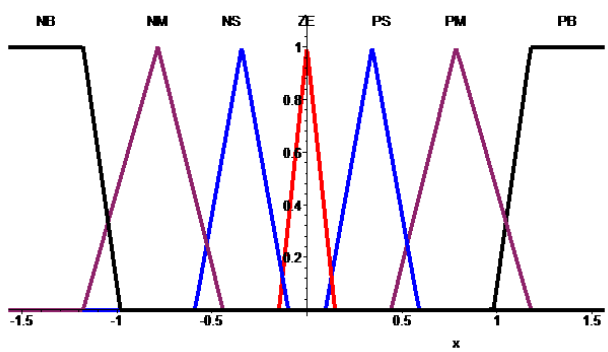


Figure 18.6.: Standard granulation with fine control at the center (the target).

consequent fuzzy set (or in the Sugeno type controllers a consequent real number).

This book has already presented two methods of using the compatibility indexes of the antecedent to construct the consequent fuzzy set. In the Mamdani method the $\min$ operator was used to truncate each consequent fuzzy set of a rule that fired. All of these truncated output fuzzy sets were then amalgamated to produce a final resultant fuzzy sets. An alternative was to use the product operator on the compatibility index and the consequent fuzzy set to produce scaled output fuzzy sets.

Given an input vector $\langle x_0, y_0 \rangle$ we construct the fuzzy set (relation) $A' \wedge B'$. This is typically an impulse fuzzy number which is one at $\langle x_0, y_0 \rangle$ and zero elsewhere. We then loop through every rule in the rulebase and calculate the individual results.

If we have a two-input—one-output controller then we loop on i and j where i goes from 1 to the granularity of X and j goes from 1 to the granularity of Y. All of the individual results are united, usually with the union operator $\max$, but as mentioned in the previous section, many fuzzy operators have been tried to model $\cup$. The result is a single fuzzy set C.

$$C = \bigcup_{i,j} C_{i,j} \tag{18.5}$$

$$= \bigcup_{i,j} \left([A' \wedge B'] \circ [A_i \wedge B_j] \to C_k \right) \tag{18.6}$$

In the Sugeno style FLC, the consequent of each implication is a real number, and we do not need to construct the fuzzy impulse number $A' \wedge B'$. Instead, we loop through each rule and determine if $\langle x_0, y_0 \rangle$ is in the support of the antecedent, if it is we the rule fires and produces a real number $c_{i,j}$ instead of a fuzzy set $C_{i,j}$. In the Sugeno system firing the rule looks like

$$c_{i,j} = [A_i(x_0) \wedge B_j(y_0)] \to c_k\,(x_0, y_0) \tag{18.7}$$

where $c_{i,j} = 0$ if $A_i(x_0) \wedge B_j(y_0) = 0$, that is $c_{i,j} = 0$ if $\langle x_0, y_0 \rangle$ is not in the support of $A_i \wedge B_j$. These real numbers are amalgamated with some function g (usually the weighted average) that will be represented symbolically by $\uplus$. The result is a control value, c, which is not a fuzzy set, and this result c does not need further processing, i.e., defuzzification.

$$c = \biguplus_{i,j} c_{i,j} \tag{18.8}$$

$$= \biguplus_{i,j} \left([A_i(x_0) \wedge B_j(y_0)] \to c_k\,(x_0, y_0) \right) \tag{18.9}$$

18.7. Defuzzification

If the design of the FLC results in a fuzzy set then the result is defuzzified. Many defuzzification methods have been tried, but the most common methods are the center of gravity, COG, and the median of maxima, MOM. However, as is typical in fuzzy sets, these names are not standard. The MOM defuzzification method has a multitude of names. The author of this book always has to find the actual equations used in other books and papers to determine precisely what defuzzification method is being applied. is an empirical study of many defuzzification methods, including some novel methods that try to balance speed and efficiency.

18.7.1. n input controllers

While both this and the previous chapter have focused on two-input—one-output FLCs a fuzzy controller can be designed with any number of input variables. There is always a single output since it is easier and faster to build many simple controllers rather than design one very complex controller.

Suppose that here are m input values and one output control value. Then the input looks like

$$\mathbf{x} = \left\langle \begin{array}{ccccccc} x_1 & x_2 & \cdots & x_i & \cdots & x_m \end{array} \right\rangle \tag{18.10}$$

where x_i is an element of the variable domain X_i.

We assume that there are n rules in the rulebase. The values of n will be the less than or equal to the granularity factors of each input variable. We assume that $A_{i,,j}$ is a fuzzy set defined on the input domain X_i and C_j is a fuzzy set defined on the output space Z. Then rule j, $\Re_j$, will look like

$$\Re_j : A_{1,j} \wedge A_{2,j} \wedge \cdots \wedge A_{i,j} \wedge \cdots \wedge A_{m,n} \to C_j \tag{18.11}$$

or

$$\Re_j : \bigwedge_i A_{i,j} \to C_j \tag{18.12}$$

The compatibility index of input $\mathbf{x}$ with rule $\Re_j$ is defined as ω_j and

$$\omega_j = A_{1,j}(x_1) \wedge A_{2,j}(x_2) \wedge \cdots \wedge A_{i,j}(x_i) \wedge \cdots \wedge A_{m,n}(x_m) = \tag{18.13}$$

$$= \bigwedge_i A_{i,j}(x_i) \; . \tag{18.14}$$

In the following section we will indicate how this expansion alters the results.

18.8. Examples

Inference and defuzzification can be best examined by looking at some of the controller designs most common in theory and practice. We will in all the following examples make the following assumptions. For the sake of simplicity and understanding the examples illustrations will always assume that there are two inputs and that two fuzzy rules have non-zero compatibility indexes.

18.8.1. Mamdani

The fuzzy implication is modelled by Mamdani's $\min$ operator and the sentence connective also is interpreted as using a logical or between the propositions and defined by the $\max$ operator Mamdani and Assilian (1975).

For an input pair $x = \langle x_1, x_2 \rangle$ that fires two rules, $n = 2$, in the rulebase we obtain

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.15}$$
$$\omega_2 = A_{2,1}(x_1) \wedge A_{2,2}(x_2) \tag{18.16}$$

The individual rule outputs are obtained by

$$C_1(z) = \omega_1 \wedge C_1(z) \tag{18.17}$$
$$C_2(z) = \omega_2 \wedge C_2(z) \tag{18.18}$$

where C_j is the consequence of $A_{1,j} \wedge A_{2,,j}$, i.e., $A_{1,j} \wedge A_{2,,j} \to C_j$ is a member of the rulebase, for $j = 1, 2$.

Then the overall system output is $C_1(z) \vee C_2(z)$.

Finally we can employ any defuzzification strategy to determine the actual control action. Mamdami's original controller used MoM.

If we have m input variables and n rules then the firing levels of the rules, ω_j, are computed using Eq.(18.13). The individual rule outputs are

$$C_j(z) = \omega_j \wedge C_j(z) \tag{18.19}$$

where C_j is the consequence of $A_{1,j} \wedge A_{2,j} \wedge \cdots \wedge A_{m,j}$, i.e., $A_{1,j} \wedge A_{2,j} \wedge \cdots \wedge A_{m,j} \to C_j$ is a member of the rulebase, for $j = 1, 2, ...n$. The overall rulebase output is the fuzzy set C

$$C(z) = \bigvee_l C'_{k_l}(z) \tag{18.20}$$
$$= \max\left[\omega_{k_l} \wedge C_{k_l}(z)\right] \tag{18.21}$$

which must be defuzzified.

Example 109. A simple illustration of a Mamdani controller.

Assume that $A_{1,j} \equiv E_j$ and $A_{2,j} \equiv \Delta E_j$ are defined in Table (18.2) and that the Sugeno controller has the rulebase as given in Table (18.4) where C_j is also defined in Table (18.2). Finally suppose that $x_1 = 0.5$ and $x_2 = -1.1$. Two rules in the rulebase will produce a nonzero compatibility index, they are:

Rule 1 If x is $A_{1,5}$ and y is $A_{2,1}$ then z is C_6

Rule 2 If x is $A_{1,5}$ and y is $A_{2,2}$ then z is C_5

Fact: x is $\langle x_1, x_2 \rangle = \langle 0.5, -1.1 \rangle$.

Consequence: z'.

For an input pair $x = \langle x_1, x_2 \rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.22}$$
$$\omega_2 = A_{1,1}(x_1) \wedge A_{2,1}(x_2)\ . \tag{18.23}$$

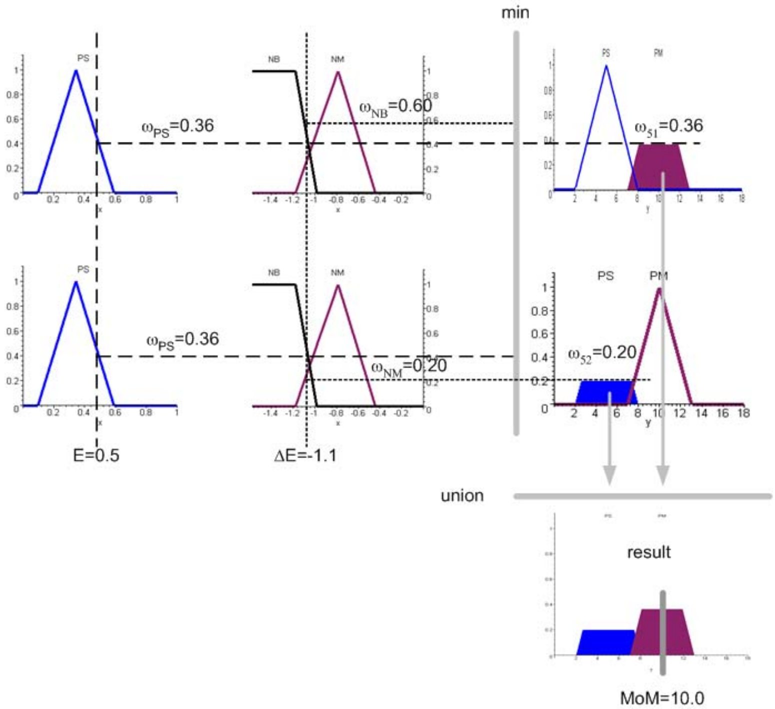


Figure 18.7.: Mamdani controller.

Then according to Fig. (18.7) we see that

$$A_{1,5}(x_1) = 0.38, \tag{18.24}$$
$$A_{2,1}(x_2) = 0.6. \tag{18.25}$$

therefore, the compatibility index of the first rule is

$$\begin{aligned} \omega_1 &= \min\left[A_{1,5}(x_1), A_{2,1}(x_2)\right] && (18.26)\\ &= \min\left[0.38, 0.6\right] && (18.27)\\ &= 0.38 && (18.28) \end{aligned}$$

and from

$$A_{1,5}(x_1) = 0.38, \tag{18.29}$$
$$A_{2,2}(x_2) = 0.2 \tag{18.30}$$

it follows that the compatibility index of the second rule is

$$\begin{aligned} \omega_2 &= \min\left[A_{1,5}(x_1), A_{2,2}(x_2)\right] && (18.31)\\ &= \min\left[0.38, 0.2\right] && (18.32)\\ &= 0.2 && (18.33) \end{aligned}$$

The fuzzy output of rule one is the fuzzy set whose membership function is

$$C_6'(z) = \omega_1 \wedge C_6(z) \tag{18.34}$$
$$= 0.38 \wedge C_6(z) \tag{18.35}$$

and the fuzzy output for rule two is the fuzzy set whose membership function is

$$C_5'(z) = \omega_2 \wedge C_5(z) \tag{18.36}$$
$$= 0.2 \wedge C_5(z) \ . \tag{18.37}$$

The final fuzzy output is the union of these two fuzzy sets,

$$C'(z) = C_6'(z) \cup C_5'(z) \tag{18.38}$$
$$= (\omega_1 \wedge C_6(z)) \vee (\omega_2 \wedge C_5(z)) \tag{18.39}$$
$$= (0.38 \wedge C_6(z)) \vee (0.2 \wedge C_5(z)) \tag{18.40}$$

and this set is illustrated in the lower right had of Fig. (18.7).

The crisp output of the fuzzy controller is arrived at by determining that the height of C' is 0.38 and that it attains this grade over the interval $[8.09, 11.91]$ which has an average value (midpoint) of

$$z' = 10 \tag{18.41}$$

so that this is the final output of the controller.

18.8.2. Larsen

In a Larsen style fuzzy controller, the compatability indices ω_j are calculated identically with the Mamdani method. At this point the Larsen and Mamdani methods diverge. In Mamdani's model the logical connective $\rightarrow$ in the fuzzy if-then statement is modeled with $\min$ and the consequent fuzzy sets end up truncated. In a Larsen fuzzy controller the logical connective $\rightarrow$ in the fuzzy if-then statement is modeled with the algebraic product t-norm and the consequent fuzzy sets are scaled instead of truncated.

For an input pair $x = \langle x_1, x_2 \rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(y) \tag{18.42}$$
$$\omega_2 = A_{1,2}(x) \wedge A_{2,2}(y) \tag{18.43}$$

The individual rule outputs are the fuzzy sets C_j obtained by

$$C_1' = \omega_1 C_1(z) \tag{18.44}$$
$$C_2' = \omega_2 C_2(z) \tag{18.45}$$

Then membership function of the inferred consequence C is pointwise given by

$$C(z) = \omega_1 C_1(z) \vee \omega_2 C_2(z). \tag{18.46}$$

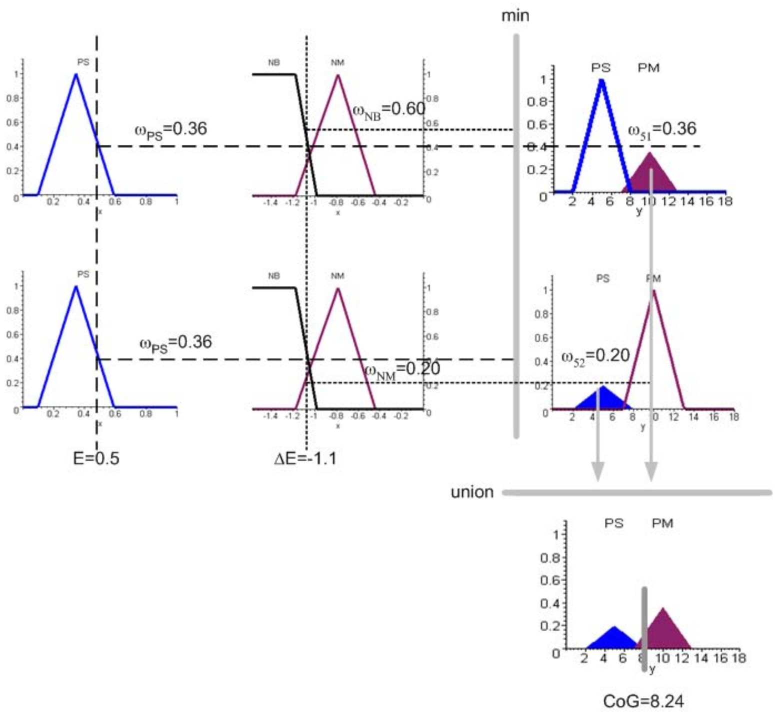


Figure 18.8.: Larsen controller.

To obtain a deterministic control action, we employ any defuzzification strategy. Usually this last step is replaced by the fast COG methodology described in Sec. (17.4).

If we have m input variables and n rules then the compatibility index of the rules, ω_j, are computed using Eq.(18.13). The individual rule outputs are

$$C'_j(z) = \omega_1 C_j(z) \tag{18.47}$$

The overall system output is the fuzzy set C

$$C(z) = \bigvee_j C'_j(z) \tag{18.48}$$

$$= \max_j \left[\omega_j C_j(z)\right] \tag{18.49}$$

which must be defuzzified.

Example 110. We illustrate Larsen's reasoning method by the following simple example:

Assume that $A_{1,j} \equiv E_j$ and $A_{2,j} \equiv \Delta E_j$ are defined in Table (18.2) and that the Larsen controller has the rulebase as given in Table (18.4) where C_j is also defined in Table (18.2). Finally suppose that $x_1 = 0.5$ and $x_2 = -1.1$. Two rules in the rulebase will produce a nonzero compatibility index, they are:

Rule 1 If x is $A_{1,5}$ and y is $A_{2,1}$ then z is C_6

Rule 2 If x is $A_{1,5}$ and y is $A_{2,2}$ then z is C_5

Fact: **x** is $\langle x_1, x_2\rangle = \langle 0.5, -1.1\rangle$.
Consequence: z'.
Fig. (18.8) shows a pictorial representation of the Larsen process.
For an input pair $x = \langle x_1, x_2\rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.50}$$
$$\omega_2 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \ . \tag{18.51}$$

The result is $\omega_1 = 0.38$ and $\omega_2 = 0.2$.
The result is the fuzzy set

$$C'(z) = \omega_1 C_6(z) \vee \omega_2 C_5(z). \tag{18.52}$$
$$= \max\left[0.38 \times C_6(z), 0.2 \times C_5(z)\right] \tag{18.53}$$

whose exact center of gravity is

$$z' = \frac{\int zC'(z)\,dz}{\int C'(z)\,dz} \tag{18.54}$$

In this case the COG is

$$z' = 8.24 \tag{18.55}$$

The fast COG algorithm notes the CoGs of C_6 is $z_1 = 10$ and of C_5 is $z_2 = 5$. The fast COG is just the weighted average of these values

$$z' = \frac{\omega_1 z_1 + \omega_2 z_2}{\omega_2 + \omega_2} \tag{18.56}$$
$$= \frac{8 \times 0.38 + 5 \times 0.2}{0.38 + 0.2} \tag{18.57}$$
$$= 8.24 \tag{18.58}$$

18.8.3. Tsukamoto

In Tsukamoto's Tsukamoto (1979) model of a fuzzy controller, the significant difference is that all the consequent fuzzy sets C_k are modeled with monotonic s-shaped membership functions. This construction ensures that these functions C_k are invertible. Thus if we know a membership grade ω in a fuzzy set C_k then there is only one domain value, z, that could have produced this value and $\omega = C_k(z)$, or in mathematical terms $z = C_k^{-1}(\omega)$.
For an input pair $x = \langle x_1, x_2\rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.59}$$
$$\omega_2 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.60}$$

In this mode of reasoning the individual crisp control actions z_1 and z_2 are computed by solving the equations

$$\omega_1 = C_1(z_1) \tag{18.61}$$
$$\omega_2 = C_2(z_2) \tag{18.62}$$

for z_1 and z_2 respectively, which gives symbolically

$$z_1 = C_1^{-1}(\omega_1) \tag{18.63}$$
$$z_2 = C_2^{-1}(\omega_2) \tag{18.64}$$

The overall crisp control action is the weighted sum (the compatibility ω_j is the weight) of the individual rule consequents z_j. The control value for two inputs is z' where

$$z' = \frac{\omega_1 z_1 + \omega_2 z_2}{\omega_2 + \omega_2} \tag{18.65}$$

If we have m input variables and n rules then the compatibility index of $\mathbf{x}$ with rule $\Re_l$, ω_l, is computed using Eq.(18.13). The crisp control action z is computed as

$$z' = \frac{\sum_{j=1}^{n} \omega_j z_j}{\sum_{j=1}^{n} \omega_j} \tag{18.66}$$
$$= \frac{\sum_{j=1}^{n} \omega_j C_j^{-1}(\omega_j)}{\sum_{j=1}^{n} \omega_j} \tag{18.67}$$

Example 111. We illustrate Tsukamoto's reasoning method by the following simple example:

Assume that $A_{1,j} \equiv E_j$ and $A_{2,j} \equiv \Delta E_j$ are defined in Table (18.2) and that the Sugeno controller has the rulebase as given in Table (18.4). In the case of the Tsukamoto controller the consequents of the firing rules C_j are functions so we will suppose that $C_6(z) = \left(\frac{z-7}{6}\right)^2$ for $7 \leq z \leq 13$ and $C_5(z) = \frac{z-2}{6}$ for $2 \leq z \leq 8$. Finally suppose that $x_1 = 0.5$ and $x_2 = -1.1$. Two rules in the rulebase will produce a nonzero compatibility index, they are:

Rule 1 If x is $A_{1,5}$ and y is $A_{2,1}$ then z is C_6

Rule 2 If x is $A_{1,5}$ and y is $A_{2,2}$ then z is C_5

Fact: $\mathbf{x}$ is $\langle x_1, x_2 \rangle = \langle 0.5, -1.1 \rangle$.

Consequence: z'.

See Fig. (18.9) for a pictorial representation of the following process.

For an input pair $x = \langle x_1, x_2 \rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.68}$$
$$\omega_2 = A_{1,1}(x_1) \wedge A_{2,1}(x_2)\ . \tag{18.69}$$

The result is $\omega_1 = 0.38$ and $\omega_2 = 0.2$.

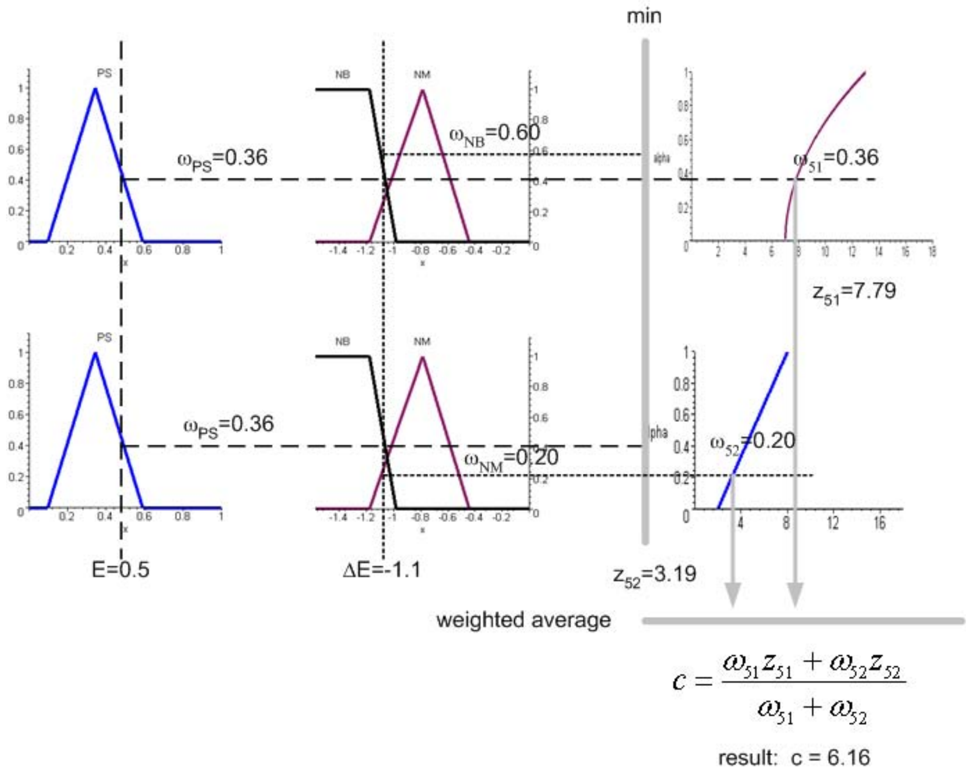


Figure 18.9.: Tsukamoto controller.

The individual output of rule one is achieved by solving

$$C_6\left(z\right) = \omega_1 \tag{18.70}$$

for z_1. Solving

$$\left(\frac{z-7}{6}\right)^2 = 0.38 \tag{18.71}$$

gives

$$z_1 = 7.79 \tag{18.72}$$

The individual output of rule two is achieved by solving

$$C_5\left(z\right) = \omega_2 \tag{18.73}$$

for z_2. Solving

$$\frac{z-2}{6} = 0.2 \tag{18.74}$$

gives

$$z_2 = 3.2 \tag{18.75}$$

The final output of the controller is the weighted sum of these z values where the weights are the compatibility indices.

The crisp control action is

$$z' = \frac{7.79 \times 0.38 + 3.2 \times 0.2}{0.38 + 0.2} \tag{18.76}$$

$$= 6.16\,. \tag{18.77}$$

18.8.4. Sugeno, Takagi, and Kang

Sugeno, Takagi and Kang (TSK) use an architecture Takagi and Sugeno (1985)Sugeno and Kang (1988) that is significantly different from the previous methods. In their system the consequents of the rules in the rule base are not fuzzy sets, instead they are equations, usually linear, that use the values of the input variable to calculate consequents.

Rule 1 If x_1 is $A_{1,1}$ and x_2 is $A_{2,1}$ then $z_1 = a_{1,1}x_1 + a_{1,2}x_2 + c_1$

Rule 2 If x_1 is $A_{2,1}$ and x_2 is $A_{2,2}$ then $z_2 = a_{2,1}x_1 + a_{2,2}x_2 + c_2$

We then fire the rule with a given input.

Fact; x is x_0 and y is y_0

Consequence: z_0

The firing levels of the rules are computed in the standard way.

$$\omega_1 = A_1(x) \wedge B_1(y) \tag{18.78}$$

$$\omega_2 = A_2(x) \wedge B_2(y) \tag{18.79}$$

The individual rule outputs are derived from the relationships

$$z_1 = a_1 x_0 + b_1 y_0 + c_1, \tag{18.80}$$

$$z_2 = a_2 x_0 + b_2 y_0 + c_2 \tag{18.81}$$

and the crisp control action z is expressed as

$$z = \frac{\omega_1 z_1 + \omega_2 z_2}{\omega_1 + \omega_2} \tag{18.82}$$

If we have m input variables and n rules then the compatibility index of $\mathbf{x}$ with rule $\Re_j$, ω_j, is computed using Eq.(18.13). Each of the n consequent z_j has the form

$$z_j = a_{1,j}x_1 + a_{2.j}x_2 + \cdots + a_{i,j}x_i + \cdots + a_{m,j}x_m + c_j \tag{18.83}$$

$$= \sum_{i=1}^{m} a_{i,j}x_i + c_j\ . \tag{18.84}$$

If we have n rules in our rule-base then the crisp control action is computed as

$$z_0 = \frac{\sum_{i=1}^{n} \omega_i z_i}{\sum_{i=1}^{n} \omega_i}\ . \tag{18.85}$$

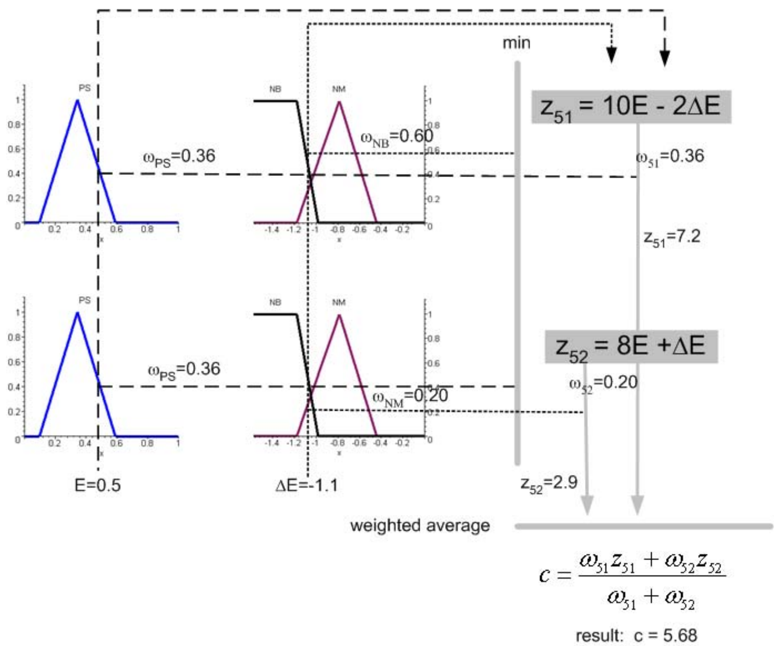


Figure 18.10.: Sugeno controller.

Example 112. We illustrate Sugeno's reasoning method by the following simple example:

Assume that $A_{1,j} \equiv E_j$ and $A_{2,j} \equiv \Delta E_j$ are defined in Table (18.2) and that the Sugeno controller has the rulebase as given in Table (18.4). In the case of the Sugeno controller the consequents of the firing rules C_j are functions of x_1 and x_2. Suppose that $C_6(x_1, x_2) = 10x_1 - 2x_2$ for and $C_5(x_1, x_2) = 8x_1 + x_2$. Finally suppose that $x_1 = 0.5$ and $x_2 = -1.1$. Two rules in the rulebase will produce a nonzero compatibility index, they are:

Rule 1 If x is $A_{1,5}$ and y is $A_{2,1}$ then $z = C_6(x_1, x_2)$.

Rule 2 If x is $A_{1,5}$ and y is $A_{2,2}$ then $z = C_5(x_1, x_2)$.

Fact: x is $\langle x_1, x_2 \rangle = \langle 0.5, -1.1 \rangle$.

Consequence: z'.

See Fig. (18.10) for a pictorial representation of the following process.

For an input pair $x = \langle x_1, x_2 \rangle$ and $n = 2$ rules in the rulebase we calculate compatibility index values ω_1 and ω_2 where

$$\omega_1 = A_{1,1}(x_1) \wedge A_{2,1}(x_2) \tag{18.86}$$
$$\omega_2 = A_{1,1}(x_1) \wedge A_{2,1}(x_2)\ . \tag{18.87}$$

The result is $\omega_1 = 0.38$ and $\omega_2 = 0.2$.

The individual output of rule one is achieved by calculating

$$z_1 = C_6(x_1, x_2) \tag{18.88}$$
$$= 10x_1 - 2x_2 \tag{18.89}$$
$$= 7.2 \tag{18.90}$$

and the output of rule two is

$$z_2 = C_5(x_1, x_2) \tag{18.91}$$
$$= 8x_1 + x_2 \tag{18.92}$$
$$= 2.9 \tag{18.93}$$

The final output of the controller is the weighted sum of these z values where the weights are the compatibility indices.

The crisp control action is

$$z' = \frac{7.2 \times 0.38 + 2.9 \times 0.2}{0.38 + 0.2} \tag{18.94}$$
$$= 5.68\,. \tag{18.95}$$

18.9. Fuzzy logic control design

1. **Granulation**
   a) Determine the input data space and the output variable space.
   b) If there are more than one output variable then divide the problem into pieces and build a separate fuzzy controller for each output variable. This reduces complexity and increases parallelism.
   c) Determine the shapes that will be used by the controller.
   d) Granulate the spaces.
2. **Fuzzification**
   a) Associate each granule with a single fuzzy set and a single linguistic term.
   b) These pieces must be sufficient to model the if and then parts of the rulebase.
3. **Rulebase**
   a) Use available experts, input/output data, and mathematical models (if they exist) to determine the rule base.
   b) Use NN, GAs, fuzzy clustering, rule of thumb, and all manner of mathematical models to determine these rules and their associated fuzzy sets.
4. **Operators**
   a) Match each logical connective in the rulebase with a fuzzy operator.
5. **Inference**

a) The rules are run in parallel.

b) You get an input value vector $\mathbf{x}_0$.

c) Mamdani

   i. Go through and find all rules where the if part of the rule has a positive membership grade for $\mathbf{x}_0$. Call this membership grade the compatibility index ω_i for rule number i.

   ii. Truncate (or scale) the then fuzzy set of the associated rule at the height ω_i.

   iii. Amalgamate the individual results with the $\max$ operator.

d) Sugeno

   i. Go through and find all rules where the if part of the rule has a positive membership grade for $\mathbf{x}_0$. Call this membership grade the compatibility index ω_i for rule number i.

   ii. Fire these rules and calculate the individual results. Call this result the control index c_i for rule number i.

   iii. Amalgamate the individual results c_i with the $weighted$ average $c = \frac{\sum_i c_i \omega_i}{\sum_i \omega_i}$.

6. **Defuzzification**

   a) Mamdani originally used what he called maximum of membership. This book uses the term median of membership.

   b) Fast center of gravity is used in most real-world applications.

18.10. Notes

Takeshi Yamakawa, in Tokyo, 1987, demonstrated the use of fuzzy control, through a set of simple dedicated fuzzy logic chips, in an “inverted pendulum” experiment. The “inverted pendulum” experiment is a classic control problem.

The original fuzzy logic controller of Mamdani Mamdani and Assilian (1975) used truncation of fuzzy sets by the compatability index and defuzzified using the mean of maxima. The Larsen Larsen (1980) controller uses scaling by the compatability index of truncation. The center of gravity method for defuzzification comes from Takagi and Sugeno Takagi and Sugeno (1985).

Tsukamoto Tsukamoto (1979), abandons fuzzy sets on the right hand side of the implications and substitutes monotone functions of the resultant compatibility index.

Sugeno, Takagi, and Kang Takagi and Sugeno (1985) Sugeno and Kang (1988) use functions on the right hand side that use all the compatibility indexes and the input data values from the *if* part of the control rules.

Homework

1. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle -0.1, .2\rangle$ if we use a Mamdani system and the Rulebase of Table 18.4.

2. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle -0.1, .2\rangle$ if we use a Larsen system and the Rulebase of Table 18.4.
3. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle -0.1, .2\rangle$ if we use a Tsukamoto system where for all $k = 1$ to 7 we set $C_k\,(z) = z^2$ and the Rulebase of Table 18.4.
4. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle -0.1, .2\rangle$ if we use a Sugeno system where for all $i = 1$ to 7 we set $C_k\,(x, y) = A_i(x)x + B_j(y)y$ and the Rulebase of Table 18.4.
5. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle 1, 1\rangle$ if we use a Mamdani system and the Rulebase of Table 18.4.
6. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle 1, 1\rangle$ if we use a Larsen system and the Rulebase of Table 18.4.
7. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle 1, 1\rangle$ if we use a Tsukamoto system where for all $k = 1$ to 7 we set $C_k\,(z) = z^2$ and the Rulebase of Table 18.4.
8. What is the control value of the Pendulum Controller upon input $\langle e, \Delta e\rangle = \langle 1, 1\rangle$ if we use a Sugeno system where for all $i = 1$ to 7 we set $C_k\,(x, y) = A_i(x)x + B_j(y)y$ and the Rulebase of Table 18.4.
9. Suggest some processes that a simple $error$ – $\Delta error$ FLC could control.
10. Design a controller to stop an elevator at the tenth floor of a building with 20 floors.
11. Find an object in your home that uses a fuzzy logic controller.

Part V.

Reference Materials

19. The Future Looks Fuzzy

A search at www.amazon.com for books on fuzzy sets in June, 2010 returned 1,909 hits. Besides control there are books applying fuzzy set theory to:

1. Social sciences
2. Image processing
3. Natural language
4. Managing uncertainty
5. Optimization
6. Economics
7. Decision making
8. Topology and algebra
9. Medicine
10. Data mining

These are just a few of the topics that one finds as one goes down the list of books on fuzzy set theory and fuzzy logic. The fundamental idea of fuzzy set theory, that we can abandon a dichotomous view of the world and instead see an infinite spectrum of variations is a momentous paradigm change similar to the paradigm change when probability was introduced. No one knows where we are headed.

A. Appendix

This chapter contains some mathematical resources that may be useful in understanding the text.

A.1. Calculus

Zeno's paradox:

> If you look at an arrow in flight, at a single instance in time the arrow is at some location, and it appears at that instant the same as a motionless arrow. Then how do we see motion?

Here is another one of those lovely paradoxes.

> Suppose we take a ruler and divide in half. Then divide the halves into halves, and keep repeating this division. If infinity exists then the process never ends, the halves eventually disappear and then a ruler is made up of pieces of nothing. Thus infinity cannot exist.

Infinity, ininitesimals, and infinite processes are the heart of calculus.

Think of the arrow as your car. When you are driving your car, the speedometer tells you how fast you are going. But at the instant you glance at the speedometer, you are at one place and velocity is distance traveled divided by time of travel. At the exact moment of the glance there is no distance just location, and there is no time duration. Thus the car cannot be moving Absurd. And even the speedometer knows better, if it is not broken. It just shows you how fast you were going over the last second and that is a pretty good estimate. If you want a better estimate, calculate the velocity over the last 10th of a second, or 100th, etc.

If you want a detailed explanation, take a calculus class. The important point is that in mathematics infinities appear and they are no problem. You can add up an infinite number of nothings to get something. This infinite sum is called an integral. You can also take an infinite series of approximations over smaller and smaller intervals and see that the approximations usually pile up on an exact answer. This infinite ratio is called a derivative.

The good thing is that both of these ideas have nice graphical interpretations.

If f is a function on the real numbers than the integral of f is the area between the graph and the x axis (see Fig (A.2)) The notation for this is

$$\int_a^b f(x)\, dx$$

which means add up all the area under the function f from where $x = a$ to where $x = b$. Integrals are primarily important because the allow us to *add up* things over

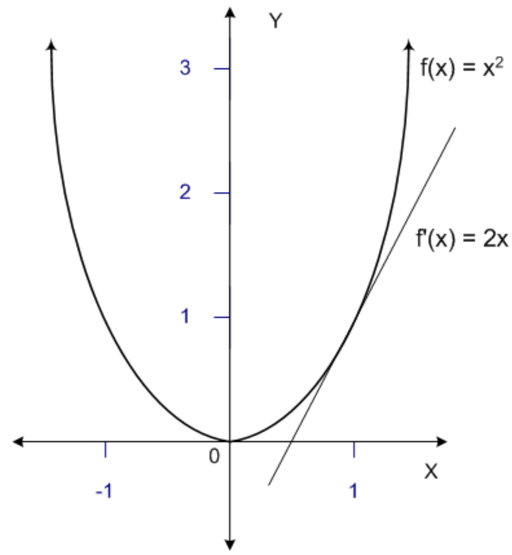


Figure A.1.: The derivative of the function f.

continuous domains like the real numbers. In addition, the integral has a nice physical meaning. If f is the *velocity* at time x then the integral is the *distance* traveled form time a to time b.

The derivative is the instantaneous rate of change. Its graphical interpretation is that it is the slope of the tangent line (see Fig. (A.1)). If the slope, the derivative, is positive then the tangent is pointing up as we go to the right and the function is increasing. If the slope is negative then the tangent is pointing down as we go to the right and the function is decreasing. In addition, the derivative has a nice physical meaning. If f is the *distance* at time x then the derivative of f is the *velocity* at time x.

There are two notations for the derivative of a function f on the real numbers. They are

$$\frac{df}{dx} \text{ and } f'(x)$$

depending on whether you like to do it the way Leibnitz or the way Newton described it.

A.2. Fuzzy Measure Theory

Fuzzy measure theory must be clearly distinguished from fuzzy set theory. While the latter is an outgrowth of classical set theory, the former is an outgrowth of classical measure theory.

The two theories may be viewed as complementary in the following sense. In fuzzy set theory, all objects of interest are precise and crisp; the issue is how much each given object is compatible with the concept represented by a given fuzzy set. In fuzzy measure theory, all considered sets are crisp, and the issue is the likelihood of

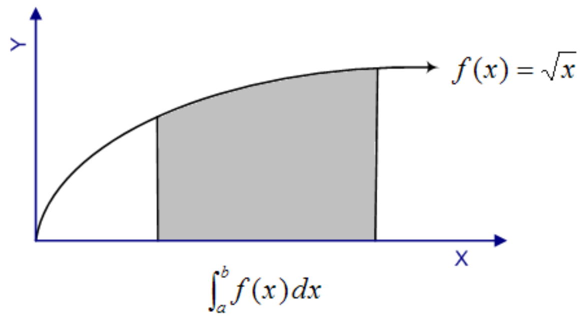


Figure A.2.: The integral of the function f.

membership in each of these sets of an object whose characterization is imprecise and, possibly, fuzzy. That is, while uncertainty in fuzzy set theory is associated with boundaries of sets, uncertainty in fuzzy measure theory is associated with boundaries of objects.

Given a universal set X and a non-empty family $\mathcal{C}$ of subsets of X (usually with an appropriate algebraic structure), a *fuzzy measure* (or *regular nonadditive measure*), g, on $\langle X, \mathcal{C}\rangle$ is a function

$$g : \mathcal{C} \rightarrow [0,1] \tag{A.1}$$

that satisfies the following requirements:

(g1) Boundary conditions — $g(\emptyset) = 0$ when $\emptyset \in \mathcal{C}$ and $g(X) = 1$ when $X \in \mathcal{C}$.

(g2) Monotonicity — for all $A, B \in \mathcal{C}$, if $A \subseteq B$, then $g(A) \leq g(B)$.

(g3) Continuity from below — for any increasing sequence $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ of sets in $\mathcal{C}$, if $\bigcup_{i=1}^{\infty} A_i \in \mathcal{C}$ then

$$\lim_{i\to\infty} g(A_i) = g(\bigcup_{i=1}^{\infty} A_i). \tag{A.2}$$

(g4) Continuity from above — for any decreasing sequence $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ of sets in $\mathcal{C}$, if $\bigcap_{i=1}^{\infty} A_i \in \mathcal{C}$ then

$$\lim_{i\to\infty} g(A_i) = g(\bigcap_{i=1}^{\infty} A_i). \tag{A.3}$$

A few remarks regarding this definition are needed. First, functions that satisfy requirements (g1), (g2) and only one of the requirements (g3) and (g4) are equally important in fuzzy set theory. If only (g3) is satisfied, the function is called a *lower semicontinuous fuzzy measure*; if only (g4) is satisfied, it is called an *upper semicontinuous fuzzy measure*. Secondly, when the universal set X is finite, requirements (g3) and (g4) are trivially satisfied and may thus be disregarded. Third, it is sometimes needed to define fuzzy measures in a more general way by extending the range of function g to the set of all nonnegative real numbers and excluding the second boundary condition $g(X) = 1$. This generalization is not applicable when fuzzy measures are utilized for characterizing uncertainty. Fourth, in this book, $\mathcal{C}$ is assumed to be a σ–algebra: $X \in \mathcal{C}$, and if $A, B \in \mathcal{C}$, then $A \cup B \in \mathcal{C}$ and $A - B \in \mathcal{C}$. In most cases, $\mathcal{C}$ is the power set, $\mathcal{P}(X)$, of X.

We can see that fuzzy measures, as defined here, are generalizations of probability measures [Billingsley, 1986] or, when conceived in the broader sense, they are generalizations of classical measures [Halmos, 1950]. In either case, the generalization is obtained by replacing the additivity requirement with the weaker requirements of monotonicity and continuity or, at least, semicontinuity. This generalization was first conceived by Sugeno [1974]. A comprehensive and up-to-date introduction to fuzzy measure theory is the subject of a graduate text by Wang and Klir [1992]. Various aspects of fuzzy measure theory are also covered in books by Denneberg [1994], Grabisch et al. [1995], and Pap [1995].

Our primary interest in this book does not involve the full scope of fuzzy measure theory, but only three of its branches: evidence theory, probability theory, and possibility theory. Relevant properties of these theories are introduced in the rest of this chapter. Fuzzy measure theory is covered here because it represents a broad, unifying framework for future research regarding uncertainty-based information.

One additional remark should be made. Fuzzy measure theory, as well as any of its branches, may be combined with fuzzy set theory. That is, function g characterizing a fuzzy measure may be defined on fuzzy sets rather than crisp sets [Wang and Klir, 1992].

A.3. Generating Functions

The next set of results exposes the relationship between t-norms, t-conorms and complements. Most t-norms, t-conorms and complements can be generated using an increasing or decreasing function as appropriate. We start with the simplest

Theorem 13 (First Characterization Theorem of Fuzzy Complements). *Let* c *be a function from* $[0,1]$ *to* $[0,1]$. *Then,* c *is an involutive fuzzy complement iff there exists a continuous function* g *from* $[0,1]$ *on* $\mathbb{R}$ *such that* $g(0) = 0$, g *is strictly increasing, and*

$$\mathsf{c}(a) = g^{-1}(g(1) - g(a)) \tag{A.4}$$

where g^{-1} *is the inverse of* g *for all* a *in* $[0,1]$.

Functions g are usually called *increasing generators*. Each function that qualifies as an increasing generator determines an involutive fuzzy complement by the equation above.

For a Standard Fuzzy Complement the increasing generator is

$$g(a) = a. \tag{A.5}$$

For the Sugeno class of complements the increasing generator is

$$g_{\lambda}(a) = \frac{1}{\lambda} ln(1 + \lambda a) \tag{A.6}$$

with $\lambda > -1$.

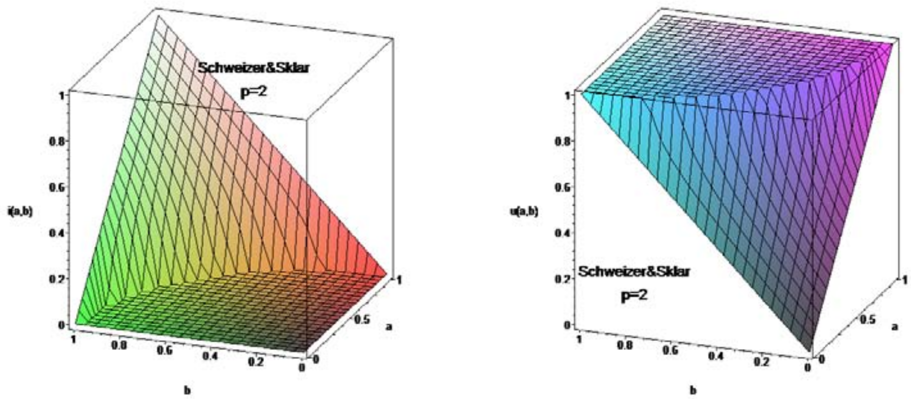


Figure A.3.: Schweizer and Sklar t-norm and t-conorm functions for $p = 2$.

For the Yager class of complements,

$$g_w(a) = a^w \tag{A.7}$$

with $w > 0$.

We can combine these to give a double parameterized increasing generator:

$$g_{\lambda,w}(a) = \frac{1}{\lambda} ln\left(1 + \lambda a^w\right) \tag{A.8}$$

with $\lambda > -1$ and $w > 0$. This yields

$$\mathsf{c}_{\lambda,w}(a) = \left(\frac{1 - a^w}{1 + \lambda a^w}\right)^{1/w} \tag{A.9}$$

which contains both the Sugeno class and the Yager class as special subclasses.

As one more example, the increasing generator

$$g_\gamma(a) = \frac{a}{\gamma + (1-\gamma)a} \tag{A.10}$$

with $\gamma > 0$ produces the class of involutive fuzzy complements

$$\mathsf{c}_\gamma(a) = \frac{\gamma^2(1-a)}{a + \gamma^2(1-a)} \tag{A.11}$$

with $\gamma > 0$.

Involutive fuzzy complements can also be produced by *decreasing generators*.

Theorem 14. *[Second Characterization Theorem of Fuzzy Complements]Let* c *be a function form* $[0, 1]$ *to* $[0, 1]$*. Then* c *is an involutive fuzzy complement iff there exists* *a continuous function f from $[0,1]$ to $\mathbb{R}$ such that $f(1)=0$, f is strictly decreasing, and $\mathsf{c}(a)=f^{-1}(f(0)-f(a))$, f^{-1} is the inverse of f for all a in $[0,1]$.*

For a Standard Fuzzy Complement the decreasing generators is,

$$f(a) = -ka + k \tag{A.12}$$

for all $k > 0$.

For the Yager class of complements the generating function is,

$$f_w(a) = 1 - a^w \tag{A.13}$$

with $w > 0$.

The *pseudo-inverse* of a decreasing generator f, denoted $f^{(-1)}$, is a function from $\mathbb{R}$ to $[0,1]$ given by

$$f^{(-1)}(a) = \begin{cases} 1 & \text{for } a \in (-\infty, 0) \\ f^{-1}(a) & \text{for } a \in [0, f(0)] \\ 0 & \text{for } a \in (f(0), \infty) \end{cases} \tag{A.14}$$

Some examples of generating functions are

$$f_1(a) = 1 - a^p \text{ with } a \in [0,1] \quad \text{and } p > 0 \tag{A.15}$$

$$f_2(a) = -\ln a \text{ with } a \in [0,1] \quad \text{and } f_2(0) = \infty. \tag{A.16}$$

These functions have pseudo-inverses

$$f_1^{(-1)}(a) = \begin{cases} 1 & \text{for } a \in (-\infty, 0) \\ (1-a)^{1/p} & \text{for } a \in [0, 1] \\ 0 & \text{for } a \in (1, \infty) \end{cases} \tag{A.17}$$

and

$$f_2^{(-1)}(a) = \begin{cases} 1 & \text{for } a \in (-\infty, 0) \\ e^{-a} & \text{for } a \in [0, \infty) \end{cases} . \tag{A.18}$$

A decreasing generator and its pseudo-inverse satisfy $f^{(-1)}(f(a)) = a$ for any a in $[0,1]$ and

$$f\left(f^{(-1)}(a)\right) = \begin{cases} 0 & \text{for } a \in (-\infty, 0) \\ a & \text{for } a \in [0, f(0)] \\ f(0) & \text{for } a \in (f(0), \infty) \end{cases} \tag{A.19}$$

The pseudo-inverse of an increasing generator g, denoted $g^{(-1)}$, is a function from $\mathbb{R}$ to $[0,1]$ defined by

$$g^{(-1)}(a) = \begin{cases} 0 & \text{for } a \in (-\infty, 0) \\ g^{-1}(a) & \text{for } a \in [0, g(1)] \\ 1 & \text{for } a \in (g(1), \infty) \end{cases} \tag{A.20}$$

Some examples are

$$g_1(a) = a^p \text{ with } a \in [0,1] \quad \text{and } p > 0 \tag{A.21}$$

$$g_2(a) = -\ln(1-a) \text{ with } a \in [0,1] \quad \text{and } g_2(1) = \infty. \tag{A.22}$$

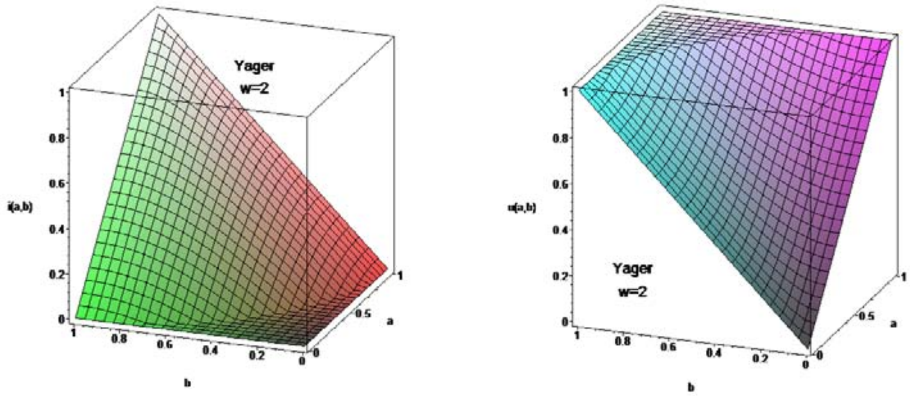


Figure A.4.: Yager t-norm and t-conorm functions for $p = 2$.

$$g1(a) = a^p (p > 0) \tag{A.23}$$

These functions have pseudo-inverses

$$g_1^{(-1)}(a) = \begin{cases} 0 & \text{for } a \in (-\infty, 0) \\ a^{1/p} & \text{for } a \in [0,1] \\ 1 & \text{for } a \in (1, \infty) \end{cases} \tag{A.24}$$

and

$$g_2^{(-1)}(a) = \begin{cases} 1 & \text{for } a \in (-\infty, 0) \\ 1 - e^{-a} & \text{for } a \in [0, \infty) \end{cases} . \tag{A.25}$$

A decreasing generator and its pseudo-inverse satisfy $g^{(-1)}(g(a)) = a$ for any a in $[0, 1]$ and

$$g\left(g^{(-1)}(a)\right) = \begin{cases} 0 & \text{for } a \in (-\infty, 0) \\ a & \text{for } a \in [0, g(1)] \\ g(1) & \text{for } a \in (g(1), \infty) \end{cases} \tag{A.26}$$

Lemma 3. *Let f by a decreasing generator. Then a function g defined by $g(a) = f(0) - f(a)$ for any a in $[0, 1]$ is an increasing generator with $g(1) = f(0)$, and its pseudo-inverse $g^{(-1)}$ is given by*

$$g^{(-1)}(a) = f^{(-1)}(f(0) - a) \tag{A.27}$$

for any a in $\mathbb{R}$.

Lemma 4. *Let g be an increasing generator. Then a function f defined by $f(a) = g(1) - g(a)$ for any a in $[0, 1]$ is a decreasing generator with $f(0) = g(1)$, and its pseudo-inverse $f^{(-1)}$ is given by*

$$f^{(-1)}(a) = g^{(-1)}(g(1) - a) \tag{A.28}$$

for any a in $\mathbb{R}$.

Theorem 15. *[Characterization Theorem of* t*-norms]Let* t *be a binary operation on the unit interval. Then,* t *is an Archimedean* t*-norm iff there exists a decreasing generator f such that*

$$\mathsf{t}(a,b) = f^{(-1)}(f(a)+f(b)) \tag{A.29}$$

for all a, b in $[0,1]$.

Theorem 16. *[Characterization Theorem of* t*-conorms]Let* s *be a binary operation on the unit interval. Then,* s *is an Archimedean* t*-conorm iff there exists an increasing generator g such that*

$$\mathsf{t}(a,b) = g^{(-1)}(g(a)+g(b)) \tag{A.30}$$

for all a, b in $[0,1]$.

Below are various examples.

Example 113. [Schweizer and Sklar]Schweizer and Sklar (1963) The class of decreasing generators parameterized by p

$$f_p(a) = 1 - a^p \tag{A.31}$$

where p is not 0 generate the pseudo-inverse

$$f_p^{(-1)}(z) = \begin{cases} 1 & \text{for } z \in (-\infty, 0) \\ (1-z)^{1/p} & \text{for } z \in [0,1] \\ 0 & \text{for } z \in (1, \infty) \end{cases}$$

This pseudo-inverse in turn generates the following t-norm

$$\begin{aligned} \mathsf{t}_p(a,b) &= f_p^{(-1)}(f_p(a)+f_p(b)) \\ &= f_p^{(-1)}(2 - a^p - b^p) \\ &= \begin{cases} (a^p + b^p - 1)^{1/p} & 2 - a^p - b^p \text{is in } [0,1] \\ 0 & \text{otherwise} \end{cases} \\ &= \max\left[0, (a^p + b^p - 1)^{1/p}\right] \end{aligned} \tag{A.32}$$

Example 114. [Yager]Yager (1980) The class of decreasing generators parameterized by w

$$f_w(a) = (1-a)^w \tag{A.33}$$

where $w > 0$ generate the pseudo-inverse

$$f_w^{(-1)}(z) = \begin{cases} 1 & \text{for } z \in (-\infty, 0) \\ 1 - z^{1/w} & \text{for } z \in [0,1] \\ 0 & \text{for } z \in (1, \infty) \end{cases} \tag{A.34}$$

This pseudo-inverse in turn generates the following t-norm

$$\begin{aligned}
\mathrm{t}_w(a,b) &= f_w^{(-1)}(f_w(a)+f_w(b)) && \text{(A.35)}\\
&= f_w^{(-1)}((1-a)^w+(1-b)^w)\\
&= \begin{cases} 1-((1-a)^w+(1-b)^w)^{1/w} & (1-a)^w+(1-b)^w \text{is in } [0,1]\\ 0 & \text{otherwise}\end{cases}\\
&= 1-mtn(1,[(1-a)^w+(1-b)^w]^{1/w})
\end{aligned}$$

Example 115. [Frank]Frank (1979) The class of decreasing generators parameterized by s

$$f_s(a) = -ln\frac{s^a-1}{s-1}$$

with$s>0$ and s is not equal to1 generate the pseudo-inverse

$$f^{(-1)}(z) = log(s)(1+(s-1)e^{-z}) \tag{A.36}$$

This pseudo-inverse in turn generates the following t-norm

$$\begin{aligned}
\mathrm{t}_s(a,b)) &= f_s^{(-1)}(f_s(a)+f_s(b)) && \text{(A.37)}\\
&= f_s^{(-1)}\left(-ln\frac{(s^a-1)(s^b-1)}{(s-1)^2}\right)\\
&= log_s\left[1+(s-1)\frac{(s^a-1)(s^b-1)}{(s-1)^2}\right]\\
&= log_s\left[1+\frac{(s^a-1)(s^b-1)}{s-1}\right]
\end{aligned}$$

The Yager class of t-norms

$$\mathrm{t}_w(a,b) = 1-\min\left[1,((1-a)^w+(1-b)^w)^{1/w}\right] \text{ with } w>0 \tag{A.38}$$

covers the full range of t-norms expressed in the following Theorem.

Theorem 17. *Let t_w denote the class of Yager t-norms defined above, then*

$$\mathrm{t}_{\min}(a,b) <= \mathrm{t}_w(a,b) <= \min(a,b) \tag{A.39}$$

For detailed proofs of all these results see Klir and Yuan (1996).

A host of t-norms have been proposed to deal with specific problems. A selection of some well-known parametric classes of t-norms are given in Table (A.1). Various procedures are now available for obtaining these and other classes of t-norms. Various experimental procedures are available to select the appropriate t-norm for a particular application Klir and Yuan (1996).

A host of t-conorms have been proposed to deal with specific problems. Some well known parameterized classes of t-conorms are given in Table A.2.

Table A.1.: Some classes of t-norms

Formula	Parameter	Originator Year
$\mathsf{t}_{\min}$		
$[\max(0, a^p + b^p - 1)]^{\frac{1}{p}}$	$p \neq 0$	Schweizer&Sklar 1983
$\dfrac{ab}{\gamma + (1-\gamma)(a+b-ab)}$	$\gamma \in (0, \infty)$	Hamacher 1978
$\log_s \left[1 + \dfrac{(s^a - 1)\left(s^b - 1\right)}{s-1}\right]$	$s \in (0, \infty)$, $s \neq 1$	Frank 1979
$1 - \min\left[1, ((1-a)^w + (1-a)^w)^{\frac{1}{w}}\right]$	$w \in (0, \infty)$	Yager 1980a
$\dfrac{ab}{\max(a, b, \alpha)}$	$\alpha \in [0, 1]$	Dubois&Prade 1980b
$\left[1 + \left[\left(\frac{1}{a} - 1\right)^\lambda + \left(\frac{1}{b} - 1\right)^\lambda\right]^{\frac{1}{\lambda}}\right]^{-1}$	$\lambda \in (0, \infty)$	Dombi 1982

Table A.2.: Some classes of t-conorms

Formula	Parameter	Originator Year
$s_{\max}$		
$1-[\max(0,(1-a)^p+(1-b)^p-1)]^{\frac{1}{p}}$	$p \neq 0$	Schweizer&Sklar 1983
$\dfrac{a+b-(\gamma-2)ab}{1+(\gamma-1)ab}$	$\gamma \in (0,\infty)$	Hamacher 1978
$1-\log_s\left[1+\dfrac{\left(s^{1-a}-1\right)\left(s^{1-b}-1\right)}{s-1}\right]$	$s \in (0,\infty)$, $s \neq 1$	Frank 1979
$\min\left[1,(a^w+b^w)^{\frac{1}{w}}\right]$	$w \in (0,\infty)$	Yager 1980a
$1-\dfrac{(1-a)(1-b)}{\max(a,b,1-\alpha)}$	$\alpha \in [0,1]$	Dubois&Prade 1980b
$\left[1+\left[\left(\frac{1}{a}-1\right)^{-\lambda}+\left(\frac{1}{b}-1\right)^{-\lambda}\right]^{-\frac{1}{\lambda}}\right]^{-1}$	$\lambda \in (0,\infty)$	Dombi 1982

Bibliography

Aczél, J. (1966). *Lectures on Functional Equations and Their Applications*. New York: Academic Press.

A.Nakamura (1999). Conflict logic with degrees. In S. K. Pal and A. Skowron (Eds.), *Rough Fuzzy Hybridization: A New Trend in Decision-Making*. Springer-Verlag.

Ashby, W. R. (1964). Constraint analysis of many-dimensional relations. *General Systems Yearbook 9*, 99–105.

Baldwin, J. F. (1979). Fuzzy logic and fuzzy reasoning. *International Journal of Man-Machine Studies 11*(4), 465–480.

Baldwin, J. F. and N. C. F. Guild (1980). Feasible algorithms for approximate reasoning using fuzzy logic. *Fuzzy Sets and Systems 3*(3), 225–251.

Baldwin, J. F. and B. W. Pilsworth (1980). Axiomatic approach to implication for approximate reasoning with fuzzy logic. *Fuzzy Sets and Systems 3*(2), 193–219.

Ball, G. and D. Hall (1965). Isodata, an iterative method of multivariate analysis and pattern classification. In *IFIPS Congress*.

Bandemer, H. and S. Gottwald (1995). *Fuzzy Sets, Fuzzy Logic Fuzzy Methods with Applications*. New York, NY: John Wiley & Sons.

Bandemer, H. and W. Nather (1992). *Fuzzy Data Analysis*. Kluwer Academic Publishers.

Bartlett, M. S. (1945). Negative probability. *Proc. Cambridge Philos. Soc. 41*, 71–73.

Belew, R. K. and L. B. Booker (Eds.) (1991). *Proceedings of the 4th International Conference on Genetic Algorithms, San Diego, CA, USA, July 1991*. Morgan Kaufmann.

Bellman, R. E. and M. Giertz (1973). On the analytic formalism of the theory of fuzzy sets. *Information Sciences 5*, 149–156.

Bellman, R. E. and L. A. Zadeh (1970). Decision making in a fuzzy environment. *Management Science 17*, 141–164.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Bezdek, J. C. (Ed.) (1985). *Analysis of Fuzzy Information*, Boca Raton, Florida. CRC Press.

Bezdek, J. C., J. Keller, R. Krisnapuram, and N. R. Pal (1999, July). *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Volume 4 of *The Handbooks of Fuzzy Sets*. Kluwer Academic Publishers.

Bezdek, J. C. and S. K. Pal (Eds.) (1992). *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. Pascataway: IEEE.

Bien, Z. and K. C. Min (Eds.) (1995). *Fuzzy Logic and Its Applications to Engineering, Information Sciences, and Intelligent Systems*, Boston. Kluwer.

Birkhoff, G. (1961). *Lattice Theory*. Providence, Rhode ISland: American Mathematical Society.

Bremermann, H. J. (1976). Pattern Recognition. In *Systems Theory in the Social Sciences*, pp. 116–159.

Burkhardt, D. G. and P. P. Bonissone (1992). Automated Fuzzy Knowledge Base Generation and Tuning. In *Proc. of the First IEEE Fuzzy Systems Conference*, pp. 179–188.

Chi, Z., H. Yan, T. P. Hamid, and R. Tizhoosh (1998). *Fuzzy Algorithms*. New Jersey: World-Scientific.

Chvátal, V. (1983a). *Linear Programming*. New York: W. H. Freeman and Company.

Chvátal, V. (1983b). *Linear Programming*. New York: W. H. Freeman and Company.

Cox, E. (1994a). *Fuzzy Thinking*. Flamingo.

Cox, E. (1994b). *The Fuzzy Systems Handbook*. Academic Press.

Dantzig, G. B. (1963). *Linear Programming and Extensions*. Princeton, New Jersey: Princeton University Press.

Dempster, A. P. (1967a). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics 38*(2), 325–339.

Dempster, A. P. (1967b). Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika 54*(3,4), 515–528.

Dietrich Leibfried, T. P. and C. Monroe (1998). Shadows and mirrors: Reconstructing quantum states of atom motion. *Physics Today 51*(4), 22–28.

Dombi, J. (1982). A general class of fuzzy operators, the De Morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems 8*(2), 149–163.

Dubois, D. and H. Prade (1980a). *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic Press.

Dubois, D. and H. Prade (1980b). New results about properties and semantics of fuzzy set–theoretic operations. In P. P. Wang and S. K. Chang (Eds.), *Fuzzy Sets: Theory and Applications to Policy Analysis and Information Systems*, New York, pp. 59–75. Plenum Press.

Dubois, D. and H. Prade (1985). An overview of aggregation operators. *Information Sciences 36*(1,2), 85–121.

Dubois, D. and H. Prade (1987). Fuzzy numbers: An overview. *Analysis of Fuzzy Information 1*, 3–39.

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

Enderton, H. B. (1977). *Elements of Set Theory*. New York: Academic Press.

Feynman, R. A. (1987). Negative probability. In *Quantum implications : essays in honour of David Bohm*, pp. 3–18.

Frank, M. J. (1979). On the simultaneous associativity of f(x, y) and x+y-f(x, y). *Aequationes Mathematicae 19*(2,3), 194–226.

Friedman, N. and J. Y. Halpern (1999). Belief revision: a critique. *Journal of Logic, Language and Information 8*(4), 401–20. Publisher: Kluwer Academic Publishers, Netherlands.

Gaines, B. R. and M. L. G. Shaw (1986). Induction of inference rules for expert systems. *Fuzzy Sets and Systems 18*(3), 315–328.

Glover, F. and M. Laguna (1998). *Tabu Search*. Kluwer Academic Publishers.

Goguen, J. A. (1979). The logic of inexact concepts. *Synthese 19*, 325–373.

Golan, J. S. (1992). *The theory of semirings*. Longman Scientific & Technical.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc.

Goonatilake, S. and S. Khebbal (Eds.) (1994). *Intelligent Hybrid Systems*, New York, NY, USA. John Wiley & Sons, Inc.

Grant, J. (2001). Churn modeling by rough set approach.

Guiasu, S. (1993a). A unitary treatment of several known measures of uncertainty induced by probability, possibility, fuzziness, plausibility, and belief. In L. V. B. Bouchon-Maunier and R. Yager (Eds.), *Uncertainty in Intelligent Systems*, pp. 322–365. Amsterdam: Elsevier Science Publishing Co.

Guiasu, S. (1993b). Weighting independent bodies of evidence. *Lecture Notes in Computer Science 747*, 168–??

Guiasu, S. (1994). Reaching a verdict by weighting evidence. In P. P. Wang (Ed.), *Advances in Fuzzy Theory and Technology*, Volume II, pp. 167–180.

Gupta, M. M., G. N. Saridis, and B. R. Gaines (Eds.) (1977). *Fuzzy Automata and Decision Processes*, New York. North–Holland.

Gupta, M. M., R. K. R. and R. R. Yager (Eds.) (1979). *Advances in Fuzzy Set Theory and Applications*, New York. North-Holland.

Halton, J. H. (1966). An interpretation of negative probabilities. *Proc. Cambridge Philos. Soc. 62*, 83–86.

Hamacher, H. (1978). Über logische verknüpfungen unscharfer aussagen und deren zugehörige bewertungs-funktionen. In *Progress in Cybernetics and Systems Research*, pp. 276–287.

Herrera, F. and J. L. Verdegay (Eds.) (1996). *Genetic Algorithms and Soft Computing*. Physica-Verlag.

Holland, J. H. (1975). *Adaptation in Natural and Artifical Systems*. Ann Arbor, MI: University of Michigan Press.

Jain, A. K., M. N. Murty, and P. J. Flynn (1999). Data clustering: a review. *ACM Computing Surveys 31*(3), 264–323.

Jang, M.-S., R. Sun, and E. Mizutani (1997). *Neuro-Fuzzy and Soft Computing*. Upper Saddle River, NJ: Prentice-Hall.

Jeffrey, R. C. (1983). *The Logic of Decision*. Chicago: University of Chicago Press.

Jones, A., A. K. and H.-J. Zimmermann (Eds.) (1986). *Fuzzy Sets Theory and Applications*, Dordrecht. D. Reidel.

Kandel, A. (1982). *Fuzzy Techniques in Pattern Recognition*. New York: John Wiley and Sons.

Kandel, A. (1986). *Fuzzy Mathematical Techniques with Applications*. Reading, MA: Addison Wesley.

Kandel, A. (Ed.) (1987). *Fuzzy Expert Systems*. Addison-Wesley.

Kaufmann, A. (1975). *Introdution to the Theory of Fuzzy Subsets - Vol. 1: Fundamental Theoretical Elements*. New York: Academic Press.

Kaufmann, A. and M. Gupta (1985). *Introduction to Fuzzy Arithmetic*. New York: Reinhold.

Kempton, W. (1984). Interview methods for eliciting fuzzy categories. *Fuzzy Sets and Systems 14*(1), 43–64.

Kim, C. J. and D. Russell (1993). Automatic Generation of Membership Function and Fuzzy Rule Using Inductive Reasoning. In *IFIS '93*, pp. 93 – 96.

Klement, E. P. and W. Slany (Eds.) (1993). *Fuzzy Logic in Artificial Intelligence: Proceedings of the 8th Austrian Artificial Intelligence Conference, FLAI 93, Linz, Austria, June 28-30, 1993*, Secaucus, NJ, USA. Springer-Verlag New York, Inc.

Klir, G. J. (1997). Fuzzy arithmetic under requisite constraints. *Fuzzy Sets and Systems 91*(2), 165–175.

Klir, G. J. and T. Folger (1988). *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, NJ: Prentice Hall.

Klir, G. J. and M. J. Wierman (1998). *Uncertainty-based information*. New York: Physica-Verlag.

Klir, G. J. and B. Yuan (1995). *Fuzzy Sets and Fuzzy Logic; Theory and Applications*. Upper Saddle River, N. Y.: Prentice Hall.

Klir, G. J. and B. Yuan (Eds.) (1996). *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh* (1 ed.). Singapore: World Scientific.

Kosko, B. (1992). *Neural Networks and Fuzzy Systems*. Englewood Cliffs: Prentice Hall.

Kosko, B. (1996). *Fuzzy Engineering*. Englewood Cliffs: Prentice Hall.

Krause, P. and D. Clark (1993). *Representing Uncertain Knowledge*. Kluwer Academic Publishers Group.

Larsen, P. M. (1980). Industrial applications of fuzzy logic control. *International Journal of Man-Machine Studies 12*(1), 3–10.

Lindley, D. V. (1987). The probability approach to the treatment of uncertainty in artificial intelligence and expert systems. *Statistical Sciences 2*, 17–24.

Lowe, D. (1998). Negative probability. Seminar.

Lowen, R. (1996). *Fuzzy Set Theory: Basic Concepts, Techniques and Bibliography*. Boston: Kluwer.

Łukasiewicz, J. (1920). O logice tròjwartościowej. *Ruch Filozoficzny 5*, 169–171. Reprinted and translated in Łukasiewicz (1970).

Łukasiewicz, J. (1963). *Elements of Mathematical Logic*. Pergomon Press.

Łukasiewicz, J. (1970). *Jan Łukasiewicz, Selected Writings*. North-Holland. Edited by L. Borowski.

Mamdani, E. (1977). Applications of fuzzy set theory to control systems: A survey. In M. M. Gupta, G. N. Saridis, and B. R. Gaines (Eds.), *Fuzzy Automata and Decision Processes*, New York, pp. 77–88. North–Holland.

Mamdani, E. and N. Baaklini (1975). Prescriptive method for deriving control policy in a fuzzy–logic controller. *Electronic Letters 11*(25,26), 625–626.

Mamdani, E. H. (1976). Application of fuzzy logic to approximate reasoning using linguistic synthesis. In *Proceedings of the sixth international symposium on Multiple-valued logic*, Logan, Utah, United States, pp. 196–202.

Mamdani, E. H. and S. Assilian (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies 7*(1), 1–13.

McClelland, J. L. and D. E. Rumelhart (Eds.) (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*. Cambridge, Mass.: MIT Press.

McNeill, F. and E. Thro (1994). *Fuzzy Logic A Practical Approach*. Morgan Kaufmann.

Mendel, J. (2001). *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Englewood Cliffs: Prentice Hall.

Menger, K. (1942). Statistical metrics. *Proc. Nat. Acad. Sci. 28*, 535–537.

Mizumoto, M. (1988). Fuzzy control under varios fuzzy reasoning methods. *Fuzzy Sets and Systems 45*, 129–151.

Moore, R. E. (1963). *Interval Analysis*. New York: Prentice-Hall, Englewood Cliffs, NJ.

Moore, R. E. (1979). *Methods and Applications of Interval Analysis*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Moore, R. E. (Ed.) (1988). *Reliability in computing: the role of interval methods in scientific computing*, New York. Academic Press.

Moral, S. (1986). Construction of a probability distribution from a fuzzy information. In *Fuzzy Sets Theory and Applications*, pp. 51–60.

Mordeson, J. N., D. S. Malik, and S.-C. Cheng (2000). *Fuzzy Mathematics in Medicine*. Englewood Cliffs: Springer Verlag.

Morgan, M. G. and M. Henrion (1990). *Uncertainty*. NY NY: Cambridge University Press.

Munkres, J. R. (1975). *Topology*. Englewood Cliffs: Prentice Hall.

Nash, S. G. and A. Sofer (1996). *Linear and Nonlinear Programming*. New York: McGraw–Hill.

Nguyen, H. T., M. Sugeno, R. Tong, and R. R. Yager (1995). *Theoretical Aspects of Fuzzy Control*. New York: John Wiley and Sons.

Nguyen, H. T. and E. A. Walker (1997). *A First Course in Fuzzy Logic*. Boca Raton, Florida: CRC Press.

Novák, V. (1989). *Fuzzy Sets and Their Applications*. Adam Hilger, Bristol, U. K.

Oblow, E. M. (1986). O-theory: A probabilistic alternative to fuzzy set theory. In B. Bouchon-Meunier and R. R. Yager (Eds.), *IPMU*, Volume 286 of *Lecture Notes in Computer Science*, pp. 111–119. Springer.

Oden, G. C. (1984). Integration of linguistic information in language comprehension. *Fuzzy Sets and Systems 14*(1), 29–41.

Ovchinnikov, S. V. (1984). Representations of transitive fuzzy relations. In *Aspects of Vagueness*, Dordrecht, pp. 105–118. Reidel.

Pawlak, Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning About Data*. Boston: Kluwer.

Pawlak, Z. (2001). Rough set theory. *KI 15*(3), 38–39.

Pawlak, Z. and A. Skowron (1994). Rough membership function. In M. F. R. E Yeager and J. Kacprzyk (Eds.), *Advances in the Dempster-Schafer of Evidence*, pp. 251–271.

Post, E. L. (1921). Introduction to a general theory of elementary propositions. *American Journal of Mathematics 43*, 163–185.

Pratt, W. K. (1978). *Digital Image Processing*. New York: John Wiley & Sons.

Procyck, T. S. and E. H. Mamdani (1979). A Linguistic Self-Organizing Process Controller. *Automatica 15*, 15–30.

Rayward-Smith, V., I. Osman, C. Reeves, and G. Smith (1996). *Modern Heuristic Search Methods*. John Wiley & Sons.

Rechenberg, I. (1965). *Cybernetic Solution Path of an Experimental Problem*. Royal Aircraft Establishment, U.K.: Ministry of Aviation.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung technischer systeme nach prinzipien der biologischen evolution*. Stuttgart, Germany: Frommann-Holzboog Verlag.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics 27*, 832–835.

Rosenfeld, A. (1975). Fuzzy graphs. See Zadeh et al. (1975), pp. 77–96.

Rumelhart, D. E. and J. L. McClelland (Eds.) (1986). *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. Cambridge, Mass.: MIT Press.

Schwefel, H.-P. (1981). *Numerical Optimization for Computer Models*. New York: Wiley Interscience.

Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*. Sixth-Generation Computer Technology. New York: Wiley Interscience.

Schweizer, B. and A. Sklar (1961). Associate functions and statistical triangle inequalities. *Publicationes Mathematicae, Debrecen 8*, 169–186.

Schweizer, B. and A. Sklar (1963). Associative functions and abstract semigroups. *Publicationes Mathematicae, Debrecen 10*, 69–81.

Schweizer, B. and A. Sklar (1983). *Probability Metric Spaces*. New York: North-Holland.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, New Jersey: Princeton University Press.

Shannon, C. (1948, July). A mathematical theory of communication. *The Bell System Technical Journal 27*(3), 379–423. From the first page: "If the base 2 is used the resulting units may be called binary digits, or more briefly, *bits*, a word suggested by J. W. Tukey.". This is the first known printed instance of the word 'bit' with the meaning of binary digit.

Smets, P. and P. Magrez (1987). Implication in fuzzy logic. *Int. J. Approximate Reasoning 1*(4), 327–347.

Smets, P., A. Mamdani, D. Dubois, and H. Prade (Eds.) (1988). *Non-Standard Logics for Automated Reasoning*, London. Academic Press.

Smithson, M. (1989). *Ignorance and Uncertainty: Emerging Paradigms*. New York: Springer-Verlag.

Smithson, M. (1990). Ignorance and disasters. *International Journal of Mass Emergencies and Disasters 8*(3), 207–235.

Sugeno, M. and G. T. Kang (1988). Structure identification of fuzzy model. *Fuzzy Sets and Systems 28*(1), 15–33.

Takagi, T. and M. Sugeno (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Systems, Man & Cybernetics 15*, 116–132.

Tizhoosh, H. R. (1997). *Fuzzy Image Processing*. New York: Springer-Verlag.

Tsukamoto, Y. (1979). An approach to fuzzy reasoning method. In R. R. M.M. Gupta and R. Yager (Eds.), *Advances in Fuzzy Set Theory and Applications*, pp. 137–149.

Turksen, I. and D. Yao (1984). Representation of connectives in fuzzy reasoning: The view through normal forms. *IEEE Transactions on Systems, Man and Cybernetics 14*, 146–151.

Turksen, I. B. (1991). Measurement of membership functions and their acquisition. *Fuzzy Sets and Systems 40*(1), 5–38.

Wang, P. P. (Ed.) (1994). *Advances in Fuzzy Theory and Technology*, Durham, N. C. Duke Univ.

Wang, P. P. (Ed.) (2001). *Computing with words*. New York: Wiley.

Wang, Z. and G. J. Klir (1992). *Fuzzy Measure Theory*. New York: Plenum Press.

Watanabe, N. (1993). Statistical Methods for Estimating Membership Functions. *Japanese Journal of Fuzzy Theory and Systems 5*(4), 589–601.

Wenstop, F. (1980). Quantitative analysis with linguistic variables. *Fuzzy Sets and Systems 4*(2), 99–115.

Yager, R. R. (1980). On a general class of fuzzy connectives. *Fuzzy Sets and Systems 4*(3), 235–242.

Yager, R. R. (1981). Some properties of fuzzy relationships. *Cybernetics and Systems 12*, 123–140.

Yager, R. R. (Ed.) (1982). *Fuzzy Set and Possibility Theory*, Oxford. Pergamon Press.

Yager, R. R. (1987). On the dempster-shafer framework and new combination rules. *Information Sciences 41*, 93–137.

Yager, R. R. (1988). Ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Systems, Man and Cybernetics 18*, 183–190.

Yager, R. R., S. Ovchinnikov, R. M. Tong, and H. T. Nguyen (Eds.) (1987a). *Fuzzy Sets and Applications-Selected Papers by L. A. Zadeh.* New York: John Wiley.

Yager, R. R., S. Ovchinnikov, R. M. Tong, and H. T. Nguyen (Eds.) (1987b). *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh*, New York. John Wiley & Sons, Inc.

Yager, R. R. and L. A. Zadeh (Eds.) (1994). *Neural Networks and Soft Computing.* New York: Van Nostrand Reinhold.

Yamato, H. (1971). Sequential estimation of a continuous probability density function and the mode. *Bulletin of Mathematical Statistics 14*, 1–12.

Yeh, R. T. and S. Y. Bang (1975). Fuzzy relations, fuzzy graphs, and their application. See Zadeh et al. (1975), pp. 125–150.

Zadeh, L. and J. Kacprzyk (Eds.) (1992). *Fuzzy Logic for the Management of Uncertainty*, New York. John Wiley and Sons.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control 8*(3), 338–353.

Zadeh, L. A. (1971). Similarity relations and fuzzy orderings. *Information Sciences 3*(2), 177–200.

Zadeh, L. A. (1972a). A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics 2*, 4–34.

Zadeh, L. A. (1972b). A rationale for fuzzy control. *J. of Dynamical Systems, Measurement, and Control 94*(1), 3–4.

Zadeh, L. A. (1973, January). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man and Cybernetics SMC-3*(1), 28–44.

Zadeh, L. A. (1975a). Calculus of fuzzy restrictions. In L. A. Zadeh, K.-S. Fu, K. Tanaka, and M. Shimura (Eds.), *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, New York, pp. 1–40. Academic Press.

Zadeh, L. A. (1975b). Fuzzy logic and approximate reasoning. *Synthese 30*(1), 407–428.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems 1*(1), 3–28.

Zadeh, L. A. (1984). A mathematical theory of evidence (book review). *AI Magazine 5*(3), 81–83.

Zadeh, L. A. (1987a). The concept of a linguistic variable and its application to approximate reasoning – I. See Yager et al. (1987b), pp. 219–269.

Zadeh, L. A. (1987b). The concept of a linguistic variable and its application to approximate reasoning – II. See Yager et al. (1987b), pp. 271–327.

Zadeh, L. A., K.-S. Fu, K. Tanaka, and M. Shimura (Eds.) (1975). *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, New York. Academic Press.

Zimmermann, H.-J. (1987). *Fuzzy Sets, decision-making and Expert Systems*. Boston: Kluwer Academic Publisher.

Zimmermann, H.-J. (Ed.) (1993, September). *First European Congress on Fuzzy and Intelligent Technologies – EUFIT'93*, Volume 2, Aachen. Elite-Foundation: Verlag der Augustinus Buchhandlung.

Zimmermann, H.-J. (1996). *Fuzzy Set Theory – And Its Applications* (3 ed.). Boston: Kluwer Academic.

Zimmermann, H.-J., L. A. Zadeh, and B. R. Gaines (Eds.) (1984). *Fuzzy Sets and Decision Analysis*, Volume 20 of *TIMS Studies in the Management Sciences*, Amsterdam. North-Holland.

Index