OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG
INF FAKULTÄT FÜR INFORMATIK

Fuzzy Data Analysis

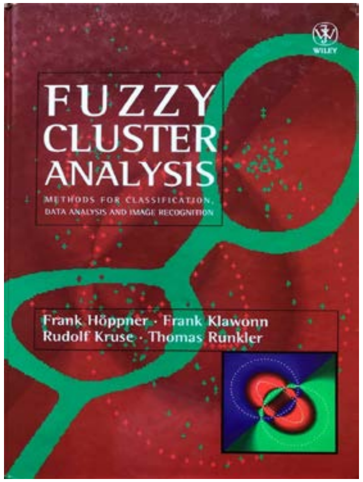
Statistics with Fuzzy Data

Prof. Dr. Rudolf Kruse

Two Interpretations of the Research on Fuzzy Data Analysis

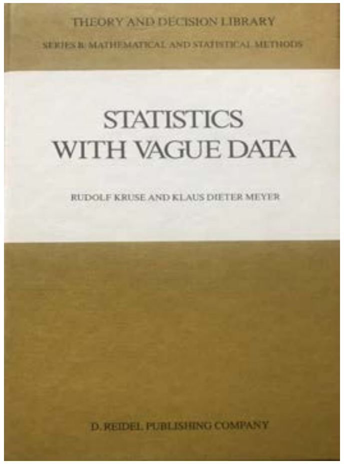
FUZZY data analysis = Fuzzy Techniques for the analysis of (crisp) data

In this course: Fuzzy Clustering



FUZZY DATA analysis = Analysis of Data that are described by Fuzzy Sets

In this course: Statistics with Fuzzy Data



Two Interpretations of Fuzzy Data

Epistemic view of fuzzy data

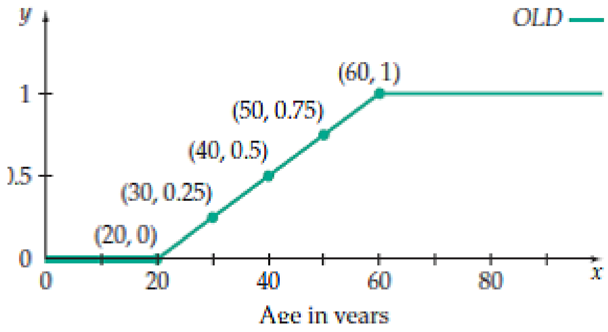
Fuzzy Sets are used to represent imcomplete knowledge about an underlying object

Example: The object is Rudolf,

a fuzzy set characterizes the knowledge

about his (unknown) age

Memberships are subjective



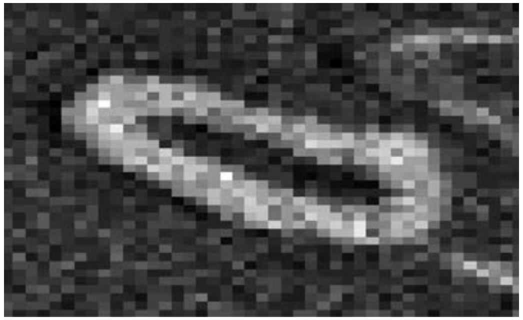
Ontic view of fuzzy data

Fuzzy Sets are considered as real, complex, graded entities

Example: The object is a photo, a fuzzy set

characterizes the content by grey level pixels

Memberships are objective, the object really exists



Epistemic Fuzzy Data

Example

Datum: “R. ate 2 or 3 eggs yesterday”

This datum is imprecise, it should be modelled by the subset {2,3 }.

Datum: “ R. ate a low number of eggs”

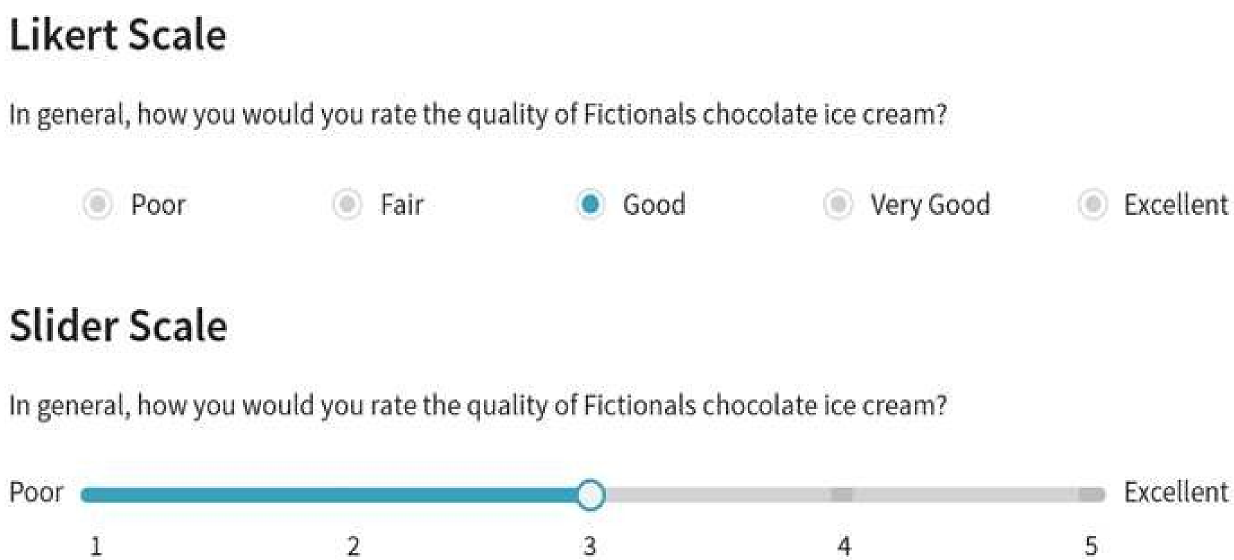
This datum is subjective, it could can be modelled by a fuzzy sets of the Natural Numbers (including 0) with an epistemic interpretation.

Crucial Question: What is the meaning of a membership degree and where do the numbers come from?

In real applications it is recommended to give a formal meaning together with a measurement method for the membership degrees values.

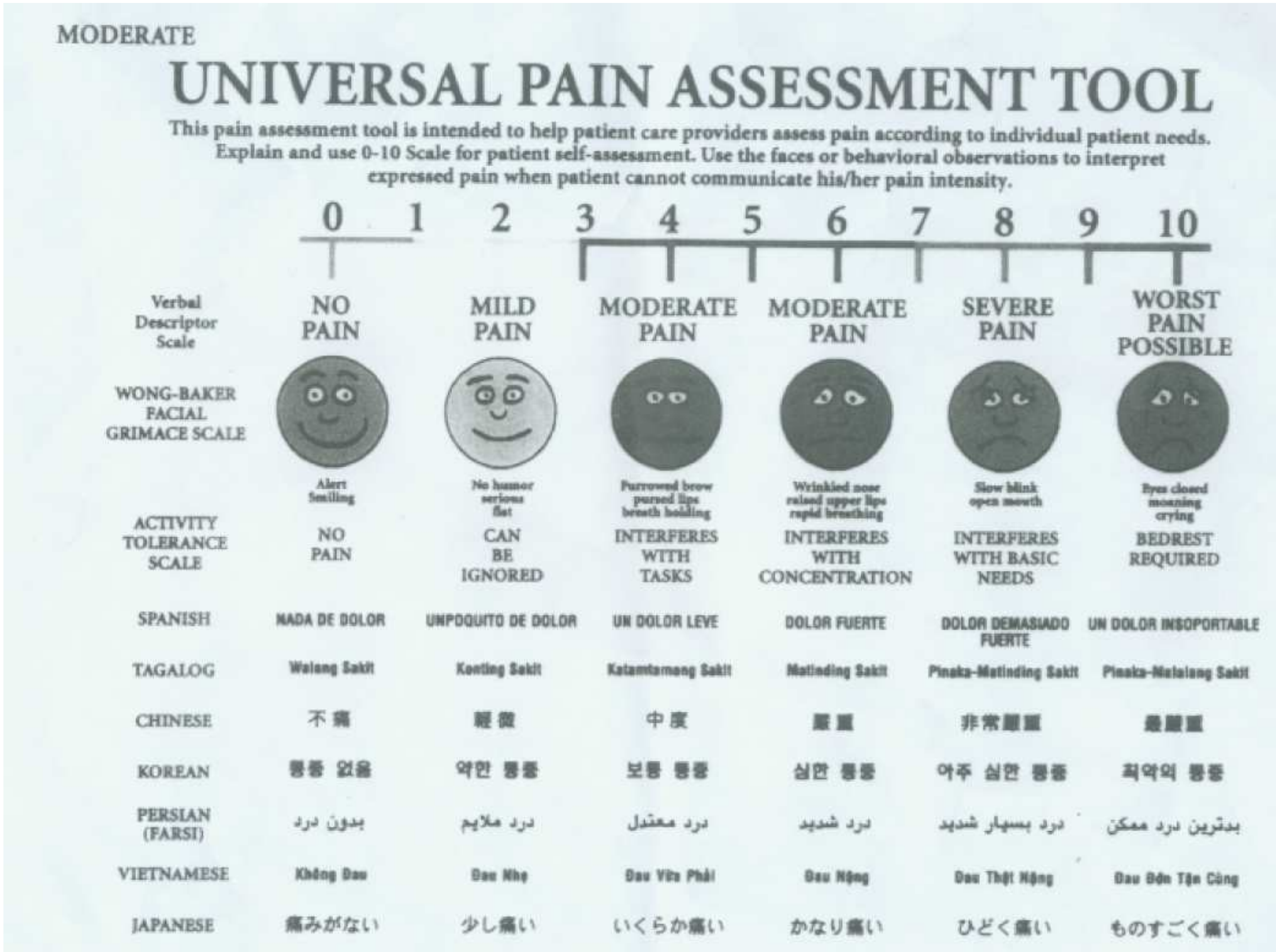
Fuzzy Membership Degrees

Gradual degrees are often used in Questionnaires:



The values of the scales can be transformed to the unit interval.

Fuzzy Membership Degrees


MODERATE

UNIVERSAL PAIN ASSESSMENT TOOL

This pain assessment tool is intended to help patient care providers assess pain according to individual patient needs. Explain and use 0-10 Scale for patient self-assessment. Use the faces or behavioral observations to interpret expressed pain when patient cannot communicate his/her pain intensity.

	0 – 1	2	3 – 4 – 5	6 – 7	8 – 9	10
Verbal Descriptor Scale	NO PAIN	MILD PAIN	MODERATE PAIN	MODERATE PAIN	SEVERE PAIN	WORST PAIN POSSIBLE
WONG-BAKER FACIAL GRIMACE SCALE	Alert Smiling	No humor serious flat	Furrowed brow pursed lips breath holding	Wrinkled nose raised upper lips rapid breathing	Slow blink open mouth	Eyes closed moaning crying
ACTIVITY TOLERANCE SCALE	NO PAIN	CAN BE IGNORED	INTERFERES WITH TASKS	INTERFERES WITH CONCENTRATION	INTERFERES WITH BASIC NEEDS	BEDREST REQUIRED
SPANISH	NADA DE DOLOR	UNPOQUITO DE DOLOR	UN DOLOR LEVE	DOLOR FUERTE	DOLOR DEMASIADO FUERTE	UN DOLOR INSOPORTABLE
TAGALOG	Walang Sakit	Konting Sakit	Katamtamang Sakit	Matinding Sakit	Pinaka-Matinding Sakit	Pinaka-Malalang Sakit
CHINESE	不痛	輕微	中度	嚴重	非常嚴重	最嚴重
KOREAN	통증 없음	약한 통증	보통 통증	심한 통증	아주 심한 통증	최악의 통증
PERSIAN (FARSI)	بدون درد	درد ملایم	درد معتدل	درد شدید	درد بسیار شدید	بدترین درد ممکن
VIETNAMESE	Không Đau	Đau Nhẹ	Đau Vừa Phải	Đau Nặng	Đau Thật Nặng	Đau Đớn Tận Cùng
JAPANESE	痛みがない	少し痛い	いくらか痛い	かなり痛い	ひどく痛い	ものすごく痛い

Fuzzy Membership Degrees

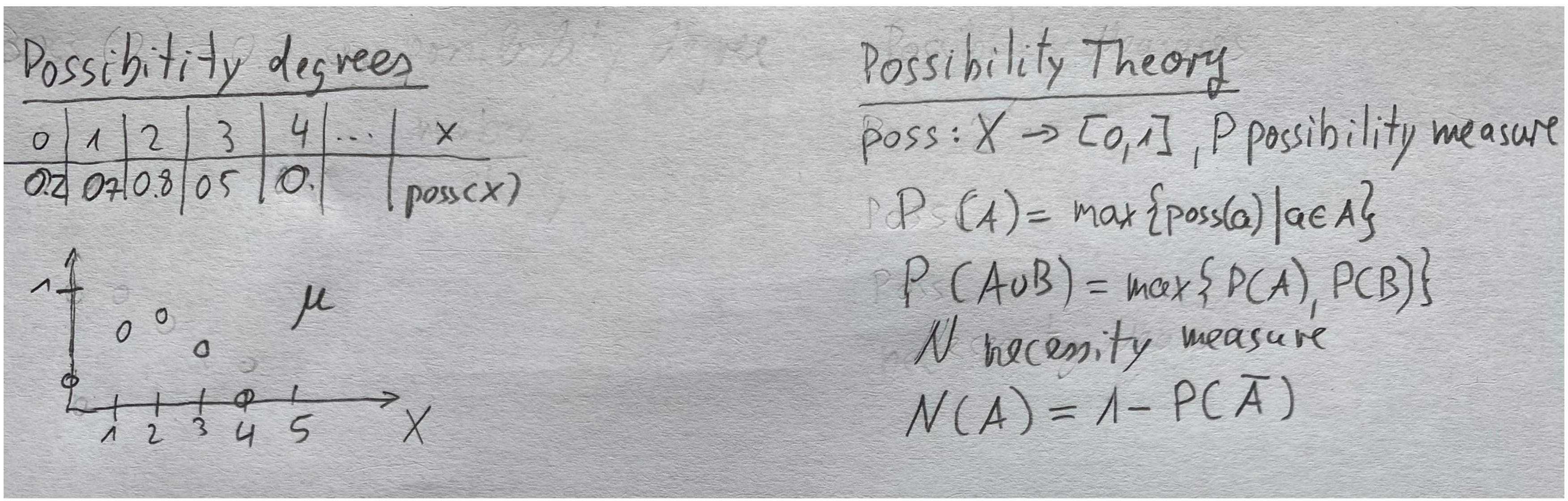
There are several different meanings of fuzzy membership degrees in real applications, often you‘ll find a

- Possibilistic Interpretation
- Preference-based Interpretation
- Frequentistic Interpretation
- Similarity-based Interpretation

Fuzzy Data based on Possibilistic Scale

Fuzzy Datum: R. ate approximately two eggs .

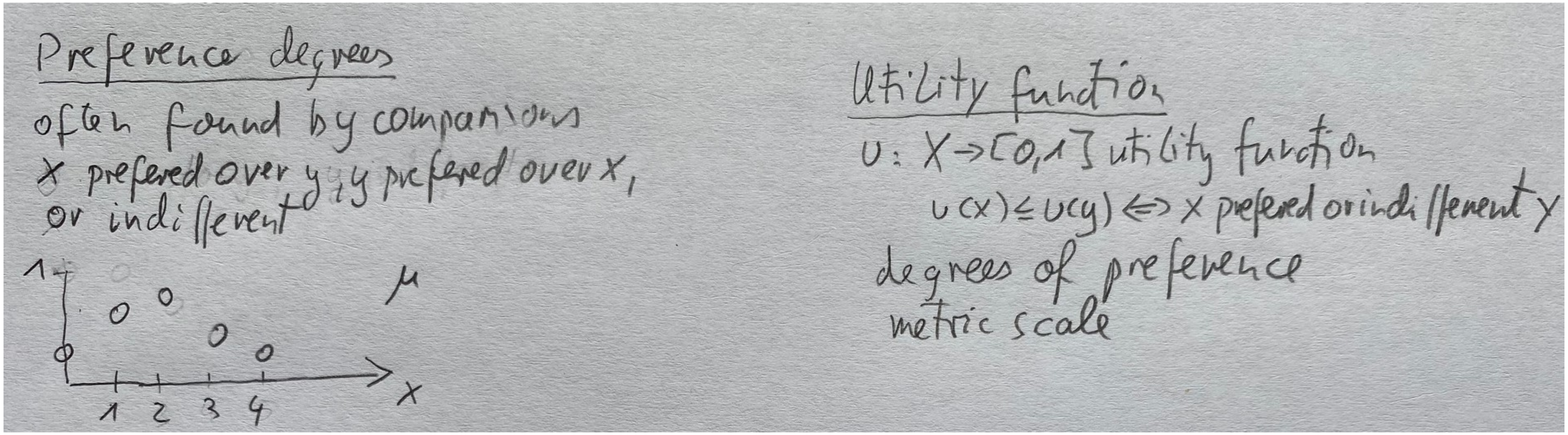
The fuzzy data are found by a reasoning process using possibilities based on physical information (how much can R. eat?) as well as epistemic information(is it possible that R. ate x eggs?



There are close links between possibility theory to Fuzzy Set Theory
In the next chapter we will study the epistemic view in more detail.

Fuzzy Data based on a Preference Scale

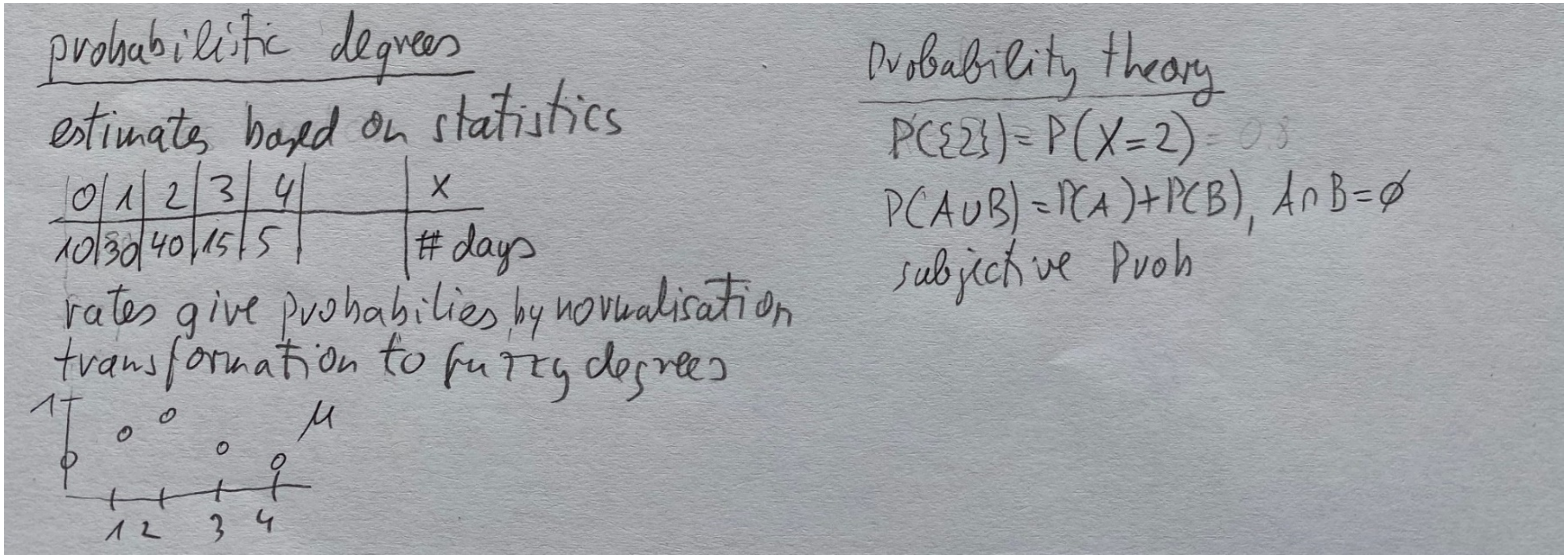
Datum: R. ate approximately two eggs.
The fuzzy data are found by a reasoning process using preferences.
How many eggs does R. prefer today?



Often this interpretation is used in optimization tasks.

Fuzzy Data based on Statistical Methods

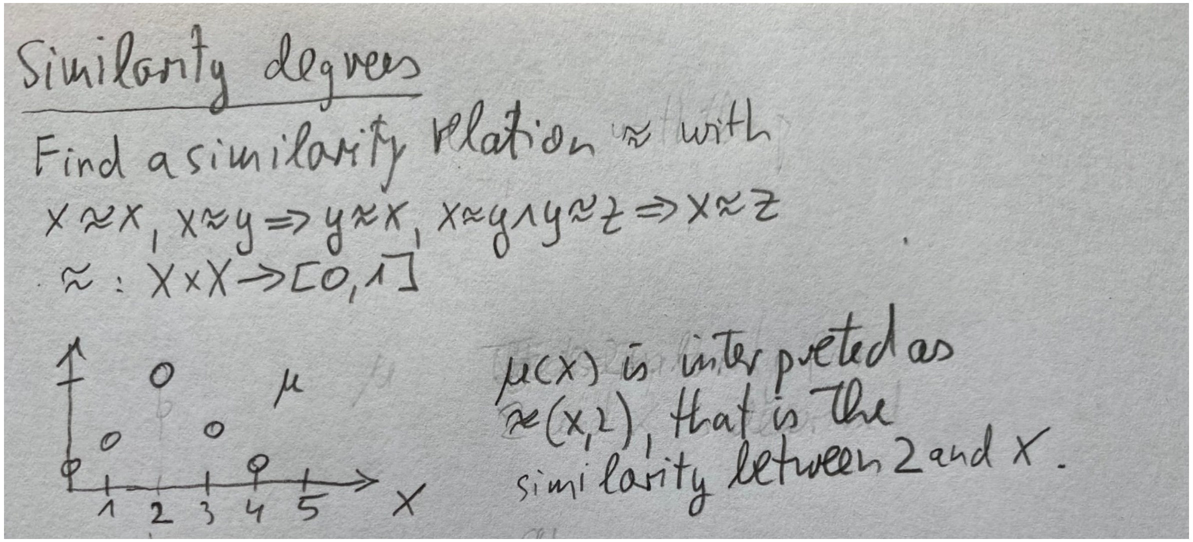
Datum: R. ate approximately two eggs.
The fuzzy data are found by a statistical analysis:
What did he eat in the last days?



Probabilistic uncertainty is modelled by statistical methods or by a version of subjective probability theory. Often the probabilities are finally transformed into a fuzzy scale.

Fuzzy Data based on similarity information

Datum: R. ate approximately two eggs.
The fuzzy data are found by similarity based (or case based) reasoning . Is a day with a similar situation for R. is known? Often the similarities are transformed into a fuzzy scale.



Similarity analysis is used in lots of other scientific disciplines.

Semantic of Fuzzy Data

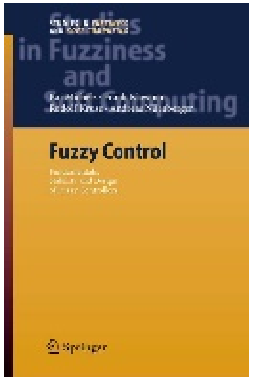
In real applications the data analysis you have to figure out (e.g. by interviews with expert), what exactly the meaning of the membership degrees is.

Different interpretations for the same fuzzy set can lead to completely different algorithms and evaluations.

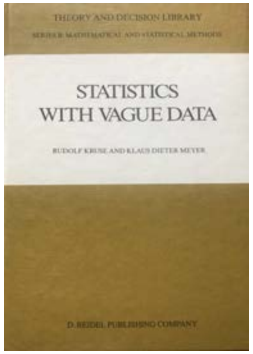
Descriptive Analysis of Imprecise Data

Experiment	Data (sample)	mean (arithmetic)	Deviation (standard)
Experiment 1	0.5, 2, 0.5	$\bar{x} = \frac{1}{n}\sum^{n} x_i = 1$	$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = 1.5$
Experiment 2 random sets, ontic view a	[0,1], {2}, [0,1]	interval arithmetik $[\frac{2}{3}, \frac{4}{3}]$	?
b	[0,1], [0,1], [0,1]	[0,1]	0
Experiment 3 random sets epistemic view a	[0,1], {2}, [0,1]	extension principle! $\{\frac{x+y+z}{3} \mid x \in [0,1], y = 2, z \in [0,1]\} = [\frac{2}{3}, \frac{4}{3}]$	$\{S(x,y,z) \mid x \in [0,1], y = 2, z \in [0,1]\}$ optimization task
b	[0,1], [0,1], [0,1]	[0,1]	$[0, \max\{S(x,y,z \mid 0 \le x,y,z \le 1\}$

Epistemic Fuzzy Data



In Fuzzy Control: **similarity based fuzzy data**



In Statistics: **possibility based fuzzy data**

How to model similarity?

Proposal 1: Similarity as Equivalence Relation?

Definition

Let A be a set and $\approx$ be a binary relation on A. $\approx$ is called an equivalence relation if and only if $\forall a, b, c \in A$,

(i)	$a \approx a$	(reflexivity)
(ii)	$a \approx b \leftrightarrow b \approx a$	(symmetry)
(iii)	$a \approx b \wedge b \approx c \rightarrow a \approx c$	(transitivity).

Let us try $a \approx b \Leftrightarrow |a - b| < \varepsilon$ where ε is fixed.

$\approx$ is not transitive, $\approx$ is no equivalence relation (Poincaré paradox)

A classical equivalence relation is not able to model similarity. What about a fuzzy version?

How to model similarity?

Proposal 2: Similarity as Fuzzy Equivalence Relation?

Definition

A function $E : X^2 \rightarrow [0, 1]$ is called a fuzzy equivalence relation with respect to a t-norm $\top$ if it satisfies for all x, y, z ∈ X the properties

(i)	$E(x, x) = 1$	(reflexivity)
(ii)	$E(x, y) = E(y, x)$	(symmetry)
(iii)	$\top(E(x, y), E(y, z)) \leq E(x, z)$	(*t*-transitivity).

$E(x, y)$ is the degree to which $x \approx y$ holds. The fuzzy relation E is also called similarity relation, *t*-equivalence relation, indistinguishability operator, or tolerance relation.

Note that property (*iii*) corresponds to the fuzzy logic statement

$$\text{if } (x \approx y) \wedge (y \approx z) \text{ then } x \approx z.$$

How to model similarity?

Proposal 3: Similarity as Lukasiewics Equivalence Relation?

Let δ be a pseudo metric on X, and let T with $\top(a, b) = \max\{a + b - 1, 0\}$ be the Łukasiewicz t-norm. Then E_δ , defined by $E_\delta(x, y) = 1 - \min\{\delta(x, y), 1\}$ is a fuzzy equivalence relation with respect to T.

For the Lukasiewicz T-norm Fuzzy equivalence and distance are dual notions. This is the only T-norm with this property.

Definition

A function $E : X^2 \to [0, 1]$ is called a (Łukasiewicz) similarity relation iff

(i) $E(x, x) = 1$ (reflexivity)

(ii) $E(x, y) = E(y, x)$ (symmetry)

(iii) $\max\{E(x, y) + E(y, z) - 1, 0\} \leq E(x, z)$ (Łukasiewicz transitivity)

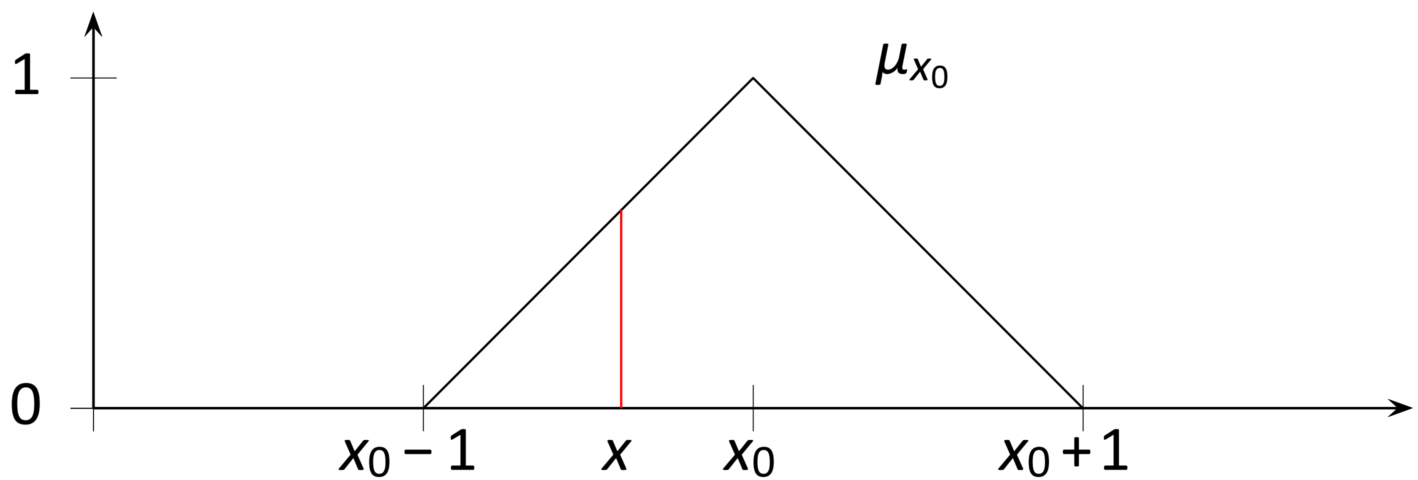
holds for all x, y, z ∈ X.

Fuzzy Set describe Local Similarity

Simple Example

$\delta(x, y) = |x - y|$ Metric

$E_\delta(x, y) = 1 - \min\{|x - y|, 1\}$ Similarity relation



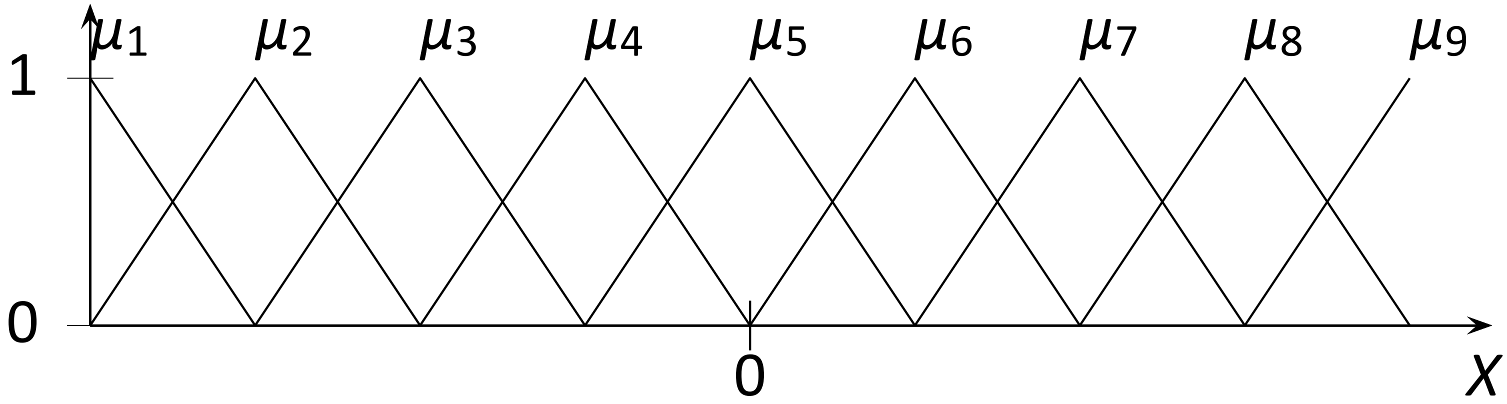
$$\mu_{x_0} : X \to [0, 1]$$

$$x \mapsto E_\delta(x, x_0)$$ Fuzzy Singleton

μ_{x_0} describes “local” similarity of points x to x_0. The membership degree is interpreted as a similarity degree.

A Fuzzy Partition models Global Similarity

Given a family of fuzzy sets that describes “local” similarities.



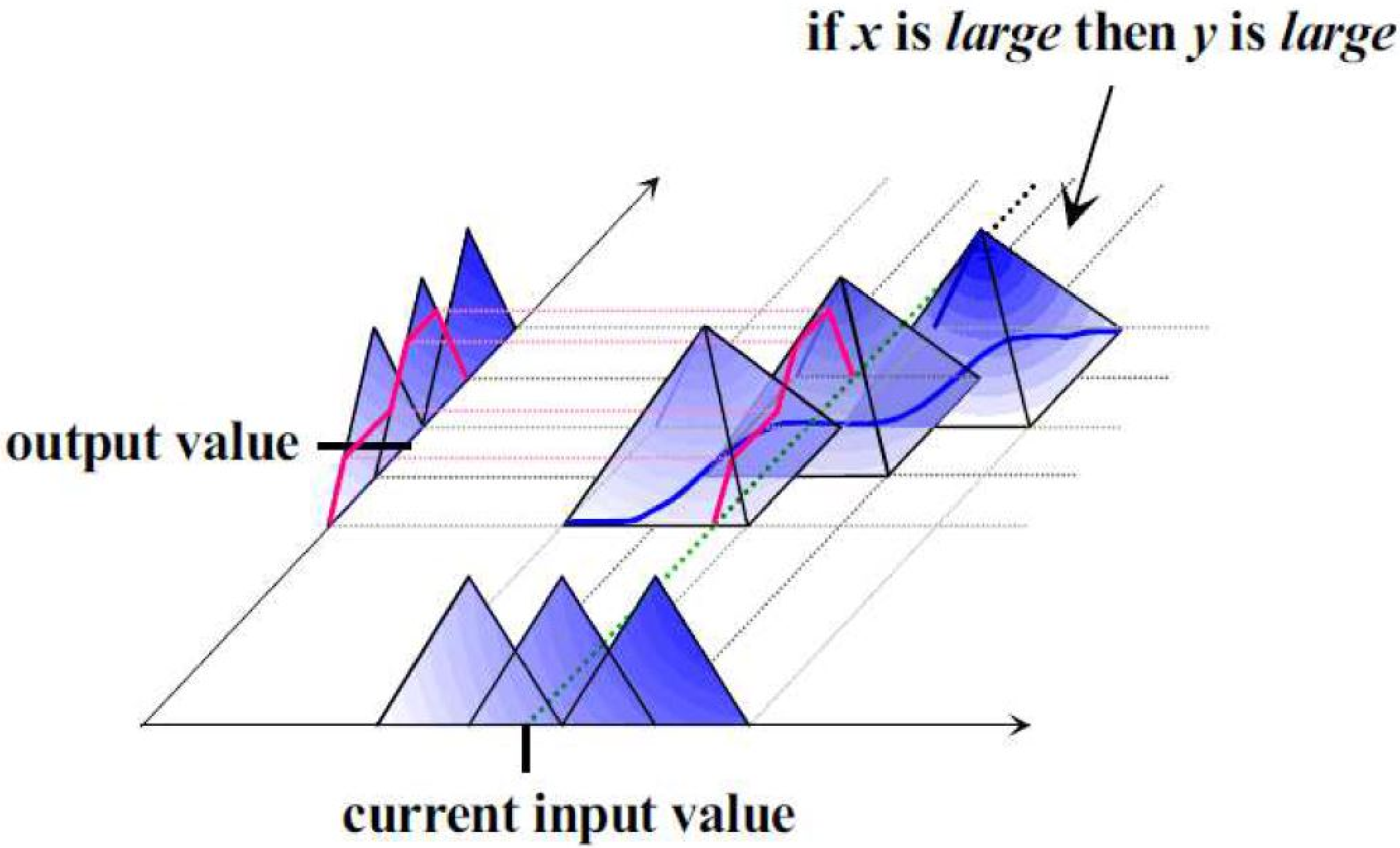
There exists a similarity relation on X with induced singletons μ_i if and only if

$$\forall i, j : \sup_{x \in X}\{\mu_i(x) + \mu_j(x) - 1\} \leq \inf_{y \in X}\{1 - |\mu_i(y) - \mu_j(y)|\}.$$

Control Engineers often have this intuitive understanding of a fuzzy datum.
Mamdani control can be seen as a similarity based interpolation.

Mamdani Control can be seen as Similarity Based Interpolation

It defines a graph that follows the pyramids

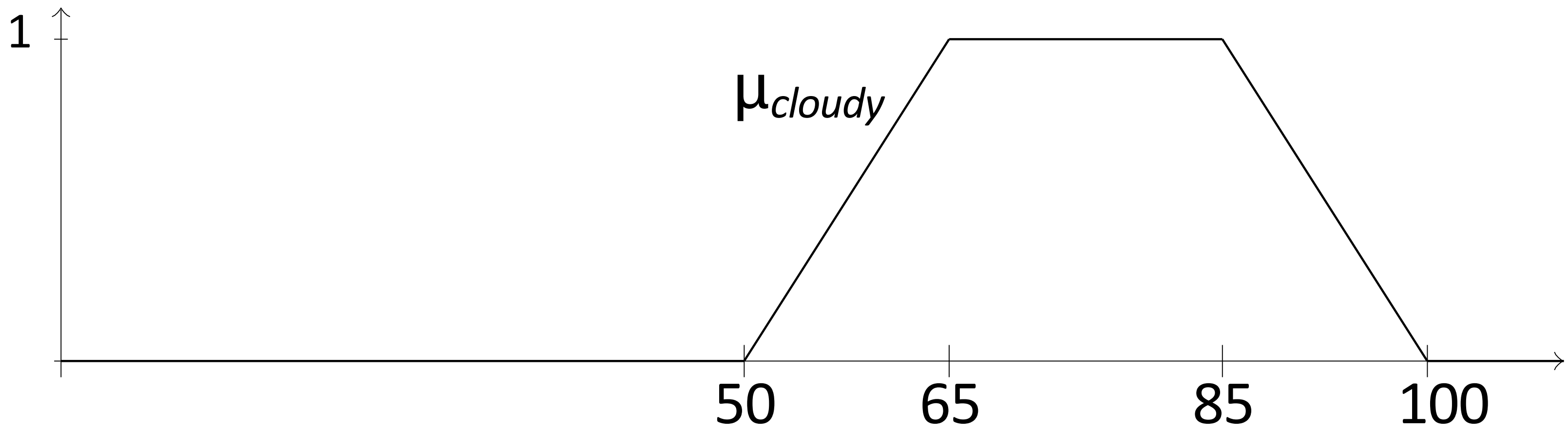


How to model possibility?

Given: It was 'cloudy', yesterday at 18, at my home

Fuzzy set $\mu_{cloudy} : X \rightarrow [0,\ 1]$, where $X = [0,\ 100]$. $x \in X$ clouding degrees in percent. The interpretation is as follows:

It exists a true clouding degree. This value is unknown, but there are additional information about the possibility of the different options.



The possibility of x is modelled by the membership degree of x: 0 means impossible, 1 totally possible, degrees between 0 and 1 indicate partial possibility.

Possibility Distributions

A function $\pi : X \rightarrow [0, 1]$ is called a possibility distribution π iff there is an $x_0 \in X$ with $\pi(x_0) = 1$.

From a mathematical point of view they are special fuzzy sets.

$\pi(u)$ is interpreted as the subjective degree of “possibility” (which is different from its probability). It quantifies a state of knowledge.

$\pi(u) = 0$: u is rejected as impossible

$\pi(u) = 1$: u is totally possible

Specificity of possibility distributions:

π is at least as specific as π' iff

for each x: $\pi(x) \leq \pi'(x)$ holds.

Possibility Measures

Let $\pi : X \to [0,1]$ a possibility distribution.

Possibility degree of $A \subseteq X$: $\Pi(A) := \sup\{\pi(x) : x \in A\}$

Necessity degree of $A \subseteq X$: $N(A) := \inf\{1 - \pi(x) : x \in \overline{A}\}$.

$\Pi(A)$ evaluates to what extent A is consistent with π

$N(A)$ evaluates to what extent A is certainly implied.

Duality expressed by: $N(A) = 1 - \Pi(\overline{A})$ for all A.

It holds:

$$\begin{aligned}
\Pi(X) &= 1 \\
\Pi(\emptyset) &= 0, \text{ and} \\
\Pi(A \cup B) &= \max\{\Pi(A), \Pi(B)\} \text{ for all } A \text{ and } B \\
\Pi(A \cap B) &\leq \min\{\Pi(A), \Pi(B)\} \text{ for all } A \text{ and } B
\end{aligned}$$

Data Analysts often have this intuitive understanding of a fuzzy datum. The true original of fuzzy datum is unknown, but there is additional information about the possibility of the different options.

Fuzzy Random Variables

Standard statistical data analysis is based on random variables $X : \Omega \rightarrow R,$

where Ω denotes the set of elementary event, and R the set of real numbers.

The concept of a random set is a generalisation. A random set $\Gamma : \Omega \rightarrow 2^R$

is a random variable where the outcomes are subsets of *R.*

The concept of a fuzzy random set is a further generalization.

A fuzzy random set $\Gamma : \Omega \rightarrow F(R)$ is a function where the outcome are fuzzy sets of R. The fuzzy sets are generated by a random mechanism.

Example: Mean Temperature

Let Ω denote the days in 2019, *P* uniform probability distribution on Ω

U(ω) Temperature on day ω at 18 h, we assume that only $T_{min}(\omega)$, $T_{max}(\omega)$ i.e. the min-max temperatures per day are recorded, but the **original** values U(ω) are unknown. We know that U(ω) is between $T_{min}(\omega)$ and $T_{max}(\omega)$, What is the mean temperature in 2019 at 18 h? We can calculate lower and upper borders by calculating the expectations of the random variables T_{min} and T_{max}.

The same method is used for handling random intervals

$\Gamma : \Omega \rightarrow 2^R$, $\Gamma(\omega) = [T_{min}(\omega), T_{max}(\omega)]$, is called a random set

$E(\Gamma) := [E(T_{min}), E(T_{max})]$ is a reasonable definition for the expected value of Γ.

This concept can be generalized from intervals to general sets.

$(\Omega, 2^{\Omega}, P)$ Probability space, a random set is a mapping $\Gamma : \Omega \to 2^{R}$

For an epistemic interpretation of the sets we define

$$E(\Gamma) := \{E(U) \mid U \text{ is random variable such that } E(U) \text{ exists and } U(\omega) \in \Gamma(\omega) \text{ for all } \omega \in \Omega \}$$

This method can be used for other quantities such as the variance.

Often subjective information about the (unknown) original data is available. In that case we can describe the data by fuzzy sets (with a possibilistic interpretation).

Using the extension principle we create a theory for descriptive statistics with fuzzy data (with an epistemic interpretation). Note that there are different models for fuzzy data with an ontic interpretation.

Expected Value of a fuzzy random variable

$\Gamma : \Omega \to \mathcal{F}(\mathbb{R})$ fuzzy random variable

The original random variable $U^*: \Omega \to \mathbb{R}$ is unknown

Given a random variable $U: \Omega \to \mathbb{R}$. we can evaluate the possibility, that U is the original, by $\inf_{\omega \in \Omega} \{(\Gamma(\omega))(U(\omega))\}$

Using the extension principle, a reasonable definition for

the expected value of Γ is obtained:

Expected value $E(\Gamma) : \mathbb{R} \to [0,1]$ fuzzy set of X:

$$x \mapsto \sup_{U:E(U)=x} \left\{ \inf_{\omega \in \Omega} \{(\Gamma(\omega))(U(\omega))\} \right\}$$

With the same method other descriptive values can be defined.