- Research Workshop -

Analysis of Uncertain Data

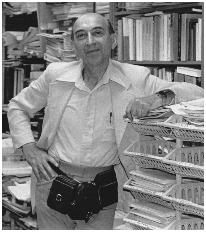
Part A: Introduction to Fuzzy Statistics

Antonio Calcagnì
University of Padova, Italy

Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska

24|26th of October, 2023

Outline



Fuzzy sets were introduced in 1965 by Lofti Zadeh (1921-2017), the father of fuzzy mathematics and fuzzy logic.

Since the seminal articles in 1965 (*Fuzzy sets* and *Fuzzy sets and systems*), research and applications based on Zadeh's work have rapidly increased.

INFORMATION AND CONTROL 8, 338–353 (1965)

Fuzzy Sets*

L. A. ZADEH

Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California

A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established. In particular, a separation theorem for convex fuzzy sets is proved without requiring that the fuzzy sets be disjoint.

FUZZY SETS AND SYSTEMS*

L. A. Zadeh

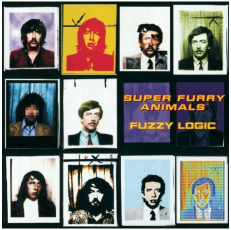
Department of Electrical Engineering, University of California, Berkeley, California

The notion of fuzziness as defined in this paper relates to situations in which the source of imprecision is not a random variable or a stochastic process, but rather a class or classes which do not possess sharply defined boundaries, e.g., the "class of bald men," or the "class of numbers which are much greater than 10," or the "class of adaptive systems," etc.

A basic concept which makes it possible to treat fuzziness in a quantitative manner is that of a fuzzy set, that is, a class in which there may be grades of membership intermediate between full membership and non-membership. Thus, a fuzzy set is characterized by a membership function which assigns to each object its grade of membership (a number lying between 0 and 1) in the fuzzy set.

Applications of fuzzy systems can be found in consumer electronics, including cameras, washes and dryers, vehicle transmission, thermostats, elevators.

The rapid spread of fuzzy set theory leads Barth Kosko to state that fuzziness is a pervasive characteristic of our reality [31].

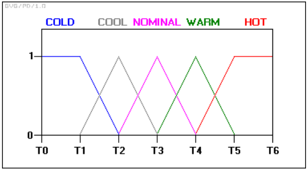


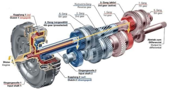
In 1996, the Welsh rock band *Super Furry Animals* released their debut album entitled *Fuzzy Logic*.

Fuzzy set theory had a big presence in our collective imagination as a revolutionary mathematical theory about fuzziness and ambiguity.

Applications of fuzzy systems can be found in:

- consumer electronics (e.g., cameras, washes, dryers)
- vehicle transmission
- thermostats, elevators
- ...





Introduction

Nowadays

Thousands of researchers working with fuzzy mathematics and related fields.

16 international associations linking researchers together.

12 international peer-reviewd journals dedicated to fuzzy theory and related disciplines.

INTERNATIONAL FUZZY SYSTEMS ASSOCIATION

EUSFLAT
EUROPEAN SOCIETY FOR FUZZY LOGIC AND TECHNOLOGY

IRAFM

POLFUZZ
POLSKIE TOWARZYSTWO ZBIORÓW ROZMYTYCH

Introduction

Fuzzy set theory is a mathematical approach dealing with problems related to ambiguousness, subjectiveness, and imprecision.

Some of the disciplines revolving around fuzzy set theory:

- fuzzy topology, geometry and algebra
- fuzzy logic
- fuzzy numerical analysis and fuzzy differential equations
- fuzzy statistics and data analysis
- control theory and intelligent systems, fuzzy optimization
- artificial intelligence, approximate reasoning
- soft computing, knowledge based systems

What is a fuzzy set?

Fuzzy sets generalize classic crisp sets

Consider the following finite reference set:

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Define the subset $A \subset \Omega$ containing *big numbers*:

$$A = \{x \in \Omega : x \in [7, 10]\}$$

In terms of characteristic function, the set A can be defined as

$$\xi_A(x) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \end{cases}$$

Then,

$$\Omega = \{1, 2, 3, 4, 5, 6, \overbrace{7, 8, 9, 10}^{A}\}$$

What is a fuzzy set?

Fuzzy sets generalize classic crisp sets

A fuzzy subset $\tilde{A} \subset \Omega$ containing *big numbers* is:

$$\tilde{A} = \{x \in \Omega : \xi_A(x) > 0\}$$

where the characteristic function represents graded levels of membership:

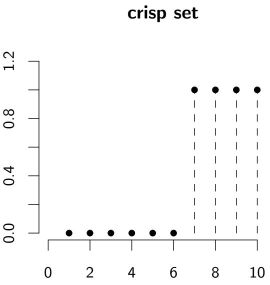
$$\xi_{\tilde{A}}(x) = \begin{cases} 0 \quad \text{if} \quad x \leq 6 \\ 0.25 \quad \text{if} \quad x = 7 \\ 0.50 \quad \text{if} \quad x = 8 \\ 0.75 \quad \text{if} \quad x = 9 \\ 1 \quad \text{if} \quad x = 10 \end{cases}$$

Then,

$$\Omega = \{1, 2, 3, 4, 5, 6, \overbrace{7, 8, 9, 10}^{\tilde{A}}\}$$

What is a fuzzy set?

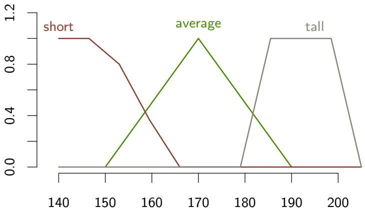
Fuzzy sets generalize classic crisp sets



Consider the simplest case of modeling the person's height:

$$\Omega = [140, 205] \subset \mathbb{R}$$

with the following fuzzy sets:



Fuzzy sets can differ in terms of shape (e.g., triangular, trapezoidal, gaussian).

What is a fuzzy set?

Fuzzy sets generalize classic crisp sets

The characteristic or membership function $\xi_{\widetilde{A}} : \Omega \rightarrow [0,1]$ provides information about the degree to which elements (or intervals) of Ω belong to $\tilde{A}$ (i.e., larger values denote higher degrees of set membership).

The use of $\xi_{\widetilde{A}}$ to represent $\tilde{A}$ is usually said **vertical representation**.

Another way to define $\tilde{A}$ is by adopting a **horizontal representation**:

$$\forall \alpha \in [0,1] : \quad [\xi_{\widetilde{A}}]_\alpha = \{x \in \Omega : \xi_{\widetilde{A}} \geq \alpha\}$$

$([\xi_{\widetilde{A}}]_\alpha)_{\alpha > 0}$ is a collection of *slices* of $\xi_{\widetilde{A}}$ called α-cuts (they are crisp sets).

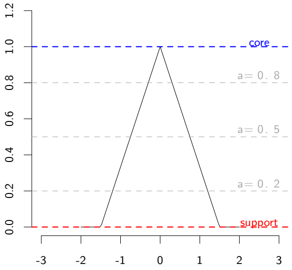
It follows that:

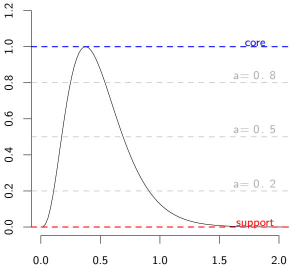
$$[\xi_{\widetilde{A}}]_{\alpha=0} = \Omega$$

$$[\xi_{\widetilde{A}}]_\alpha \supseteq [\xi_{\widetilde{A}}]_{\alpha'} \quad \alpha < \alpha'$$

What is a fuzzy set?

Fuzzy sets generalize classic crisp sets



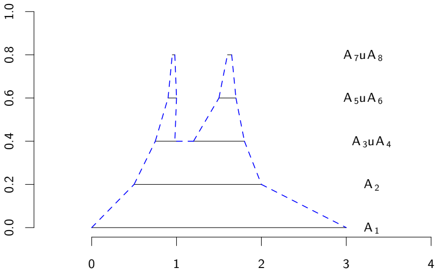


Support of $\tilde{A}$: $[\xi_{\widetilde{A}}]_{\alpha>0}$

Core of $\tilde{A}$: $[\xi_{\widetilde{A}}]_{\alpha=1}$

What is a fuzzy set?

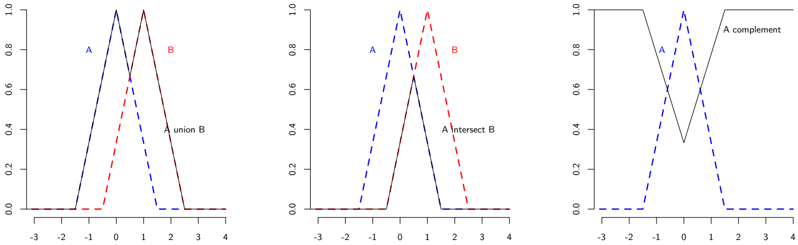
Fuzzy sets generalize classic crisp sets



The vertical representation can also be obtained via the horizontal one: $\xi_{\widetilde{A}}$ is the **upper envelope** of $([\xi_{\widetilde{A}}]_\alpha)_{\alpha \in \mathcal{L}}$.

What is a fuzzy set?

Basic operations between fuzzy sets



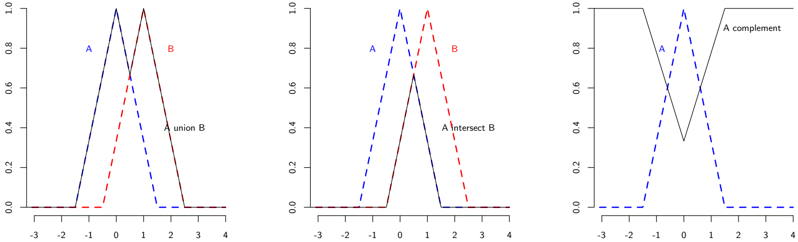
Union $\xi_{\tilde{A}\cup\tilde{B}}(x) = \max\{\xi_{\widetilde{A}}(x), \xi_{\widetilde{B}}(x)\}$

Intersection $\xi_{\tilde{A}\cap\tilde{B}}(x) = \min\{\xi_{\widetilde{A}}(x), \xi_{\widetilde{B}}(x)\}$

Complement $\overline{\xi}_{\tilde{A}}(x) = 1 - \xi_{\widetilde{A}}(x)$

What is a fuzzy set?

Basic operations between fuzzy sets



<u>Note</u>: The operations of intersection and union can also be defined in terms of **T-norm** (eg: product, Lukasiewicz) and **T-conorm** (eg: maximum, bounded sum).

Among the various type of fuzzy sets, those which are defined over $\mathbb{R}$ are of particular importance.

A fuzzy set $\tilde{X}$ is a **fuzzy number** $\tilde{x}$ if it satisfies the following conditions:

- $\tilde{X}$ is normal (i.e., its maximum degree of membership is one)
- $\tilde{X}$ is convex
- $\xi_{\widetilde{X}}$ is at least piecewise continuous

A fuzzy number $\tilde{x}$ is *positive* if $[\xi_{\widetilde{x}}]_{\alpha=0} \subseteq (0,\infty)$, *negative* if $[\xi_{\widetilde{x}}]_{\alpha=0} \subseteq (-\infty,0)$.

The set of fuzzy numbers is denoted by $\mathcal{F}(\mathbb{R})$.

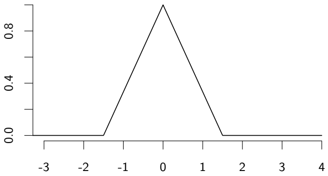
What is a fuzzy set?

Fuzzy numbers as special fuzzy sets

Some of the most common fuzzy numbers are the following:

Triangular

$$\text{tfn}(x; x_0, x_l, x_u) = \begin{cases} 0, & \text{if } x < x_l \text{ or } x > x_u \\ 1, & \text{if } x = x_0 \\ \frac{(x_0 - x)}{(x_0 - x_l)} & \text{if } x \in [x_l, x_0) \\ \frac{(x - x_0)}{(x_r - x_0)} & \text{if } x \in (x_0, x_l] \end{cases}$$



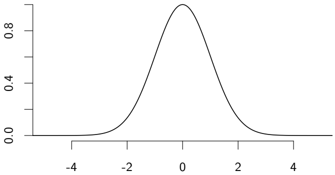
What is a fuzzy set?

Fuzzy numbers as special fuzzy sets

Some of the most common fuzzy numbers are the following:

Gaussian

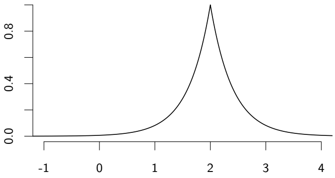
$$\text{gfn}(x; x_0, s_l, s_r) = \begin{cases} \exp\left(-\frac{1}{2s_l^2}(x - x_0)^2\right) & \text{if } x \in (-\infty, x_0) \\ \exp\left(-\frac{1}{2s_r^2}(x - x_0)^2\right) & \text{if } x \in [x_0, \infty) \end{cases}$$



Some of the most common fuzzy numbers are the following:

Exponential

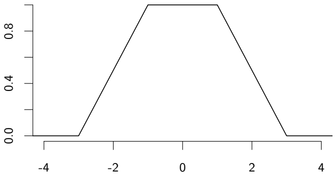
$$\text{efn}(x; x_0, t_l, t_r) = \begin{cases} \exp\left(-\frac{1}{t_l}(x - x_0)\right) & \text{if } x \in (-\infty, x_0) \\ \exp\left(-\frac{1}{t_r}(x - x_0)\right) & \text{if } x \in [x_0, \infty) \end{cases}$$



Some of the most common fuzzy numbers are the following:

Trapezoidal

$$\text{tfn}(x; x_0, x_1 x_l, x_r) = \begin{cases} 0, & \text{if } x < x_l \text{ or } x > x_u \\ 1, & \text{if } x \in [x_0, x_1] \\ \frac{(x_0 - x)}{(x_0 - x_l)} & \text{if } x \in [x_l, x_0) \\ \frac{(x - x_0)}{(x_r - x_0)} & \text{if } x \in (x_0, x_l] \end{cases}$$



Calculus with fuzzy numbers

The extension principle

To extend the most elementary operations (E) of **addition/subtraction**, **multiplication**, and **division** to handle with fuzzy numbers, the *Zadeh's extension principle* (EP) could be applied in this case:

$$\xi_{\widetilde{c}}(z) \quad = \sup_{z=E(x,y)} \min\{\xi_{\widetilde{a}}(x), \xi_{\widetilde{b}}(y)\}$$

Although it produces reasonable results in most applications, this generalization does not guarantee that $\xi_{\widetilde{c}}$ is still a fuzzy number.

Alternatives exist and they are based on generalization of the EP, use of shape/reference functions, and reduction to interval calculus.

Dubois and Prade's **LR fuzzy numbers** consist of re-parameterizing $\xi_{\tilde{c}}$ in terms of two monotonic decreasing and left-continuous *shape functions*:

$$L : \mathbb{R}^+ \rightarrow [0,1] \qquad R : \mathbb{R}^+ \rightarrow [0,1]$$

with

$$L/R(\upsilon) \begin{cases} = 0 & \text{if } \upsilon = 1 \\ = 1 & \text{if } \upsilon = 0 \\ > 0 & \text{if } \upsilon < 1 \\ < 1 & \text{if } \upsilon > 0 \end{cases}$$

and where

$$\xi_{\tilde{c}}(x) = \begin{cases} L\left(\frac{m-x}{l}\right) & \text{if } x < m \\ R\left(\frac{x-m}{r}\right) & \text{if } x \geq m \end{cases}$$

with m, l, r being the **mode**, **left/right spread** ($l > 0$,$r > 0$).

For instance, with the LR parametrization the **triangular fuzzy number**:

$$\xi_{\tilde{c}}(x) = \begin{cases} L\left(\frac{x_0 - x}{x_0 - x_l}\right) & \text{if } x < x_0 \\ R\left(\frac{x - x_0}{x_u - x_0}\right) & \text{if } x \geq x_0 \end{cases}$$

with

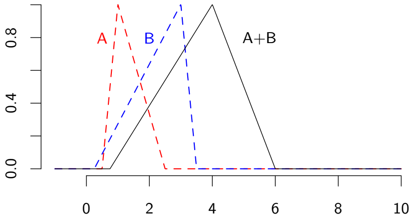
$$L(u) = R(u) = \max\{0, 1 - u\}$$

Given $\tilde{a}$ and $\tilde{b}$, in order to compute $\tilde{c} = E(\tilde{a}, \tilde{b})$ - with $E()$ being one of the basic operation - it is needed that $E()$ still produces LR-fuzzy number (i.e., *closure*). This requires some restrictions on the type of L/R shape functions being involved.

To ensures closureness of $E()$, L/R functions need to be approximated in some cases (e.g., via secant or tangent techniques). This is especially valid for *multiplication* of fuzzy numbers [27].

Addition (triangular case)

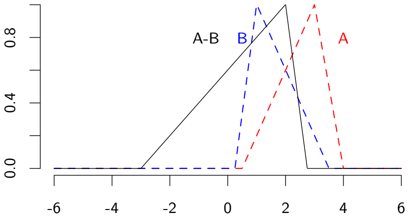
$$\tilde{c} = \tilde{a} + \tilde{b}$$
$$(m_a + m_b, l_a + l_b, r_a + r_b)_{LR} = (m_a, l_a, r_a)_{LR} + (m_b, l_b, r_b)_{LR}$$



Subtraction (triangular case)

$$\tilde{c} = \tilde{a} - \tilde{b}$$

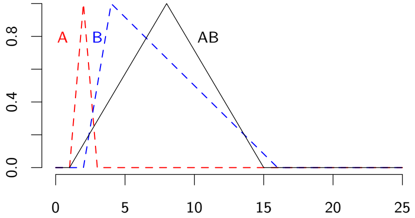
$$(m_a - m_b, l_a + r_b, r_a + l_b)_{LR} = (m_a, l_a, r_a)_{LR} - (m_b, l_b, r_b)_{LR}$$



Multiplication (triangular positive case)

$$\tilde{c} = \tilde{a} \cdot \tilde{b}$$

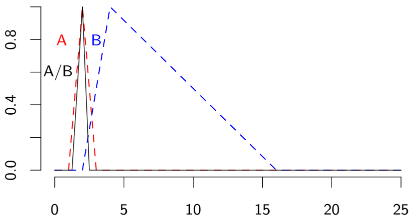
$$(m_a m_b, \underbrace{m_a l_b + m_b l_a - l_a l_b, m_a r_b + m_b r_a + r_a r_b}_{\text{secant approximation}})_{LR} = (m_a, l_a, r_a)_{LR} \cdot (m_b, l_b, r_b)_{LR}$$



Division (triangular positive case)

$$\tilde{c} = \tilde{a} \cdot \tilde{b}^{-1}$$

$$(m_a/m_b, \underbrace{(x_a r_b + x_b l_a)m_b^{-2}, (x_a l_b + x_b r_a)m_b^{-2}}_{\text{tangent approximation}})_{LR} = (m_a, l_a, r_a)_{LR}/(m_b, l_b, r_b)_{LR}$$



To overcome the difficulties of the calculus based on L-R shape functions, one can turn to the decomposition of a $\tilde{x}$ as a (finite) sequence of α intervals $([\xi_{\tilde{x}}]_\alpha)_{\alpha \in \mathcal{L}}$. In this case,

$$\begin{aligned}
\tilde{c} &= E(\tilde{a}, \tilde{b}) \\
C_\alpha &= E(A_\alpha, B_\alpha) \\
[l_a, r_a]_\alpha &= E([l_a, r_a]_\alpha, [r_a, r_b]_\alpha)
\end{aligned}$$

and the operation $E()$ is applied element-wise on the α-cuts of the fuzzy numbers being involved.

Calculus with fuzzy numbers

Interval calculus

Input: $\tilde{a}$, $\tilde{b}$ Output: $\tilde{c} = E(\tilde{a}, \tilde{b})$ using α-cuts of inputs

- **addition**

$$[l_a + r_a, l_b + r_b]_\alpha = [l_a, r_a]_\alpha + [l_b, r_b]_\alpha$$

- **subtraction**

$$[l_a - r_b, l_b - r_a]_\alpha = [l_a, r_a]_\alpha - [l_b, r_b]_\alpha$$

- **multiplication**

$$[\min(S_\alpha), \max(S_\alpha)]_\alpha = [l_a, r_a]_\alpha \cdot [l_b, r_b]_\alpha$$
$$S_\alpha = \{l_a l_b, l_a r_b, r_a l_b, r_a r_b\}$$

- **division**

$$[\min(S_\alpha), \max(S_\alpha)]_\alpha = [l_a, r_a]_\alpha / \ [l_b, r_b]_\alpha$$
$$S_\alpha = \{l_a/l_b, l_a/r_b, r_a/l_b, r_a/r_b\}$$
$$0 \notin [l_b, r_b]_\alpha$$

In some cases, the addition/subtraction property does not hold here [11]:

$$A_\alpha + (-1 \cdot B_\alpha) + B_\alpha \neq A_\alpha$$

Thus, $(\mathcal{F}(\mathbb{R}), +, \cdot)$ is a **semilinear space**.

The *lack of subtraction* requires the generalization of the classical differentiation that fit the semilinear environment. The most successful approach implies the use of **Hukuhara differentiation** (e.g., see [19]):

$$C_\alpha = A_\alpha -_H B_\alpha \quad \text{such that } A_\alpha + B_\alpha = C_\alpha$$

Note that $-_H$ does not always exist (but if it does it is unique).

The semilinear space $(\mathcal{F}(\mathbb{R}), +, \cdot)$ can be endowed with a **metric** which can be useful in statistical applications:

$$D_\tau^\lambda(A_\alpha, B_\alpha) = \left(\int_0^1 \left((\text{mid}A_\alpha - \text{mid}B_\alpha)^2 + \tau(\text{spr}A_\alpha - \text{spr}B_\alpha)^2 \right) \, d\lambda(\alpha) \right)^{\frac{1}{2}}$$

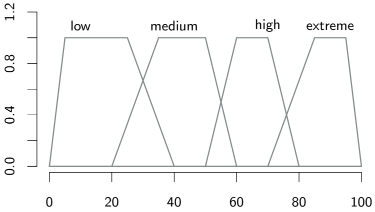
where $\text{mid}X_\alpha = (l_{x_\alpha} - r_{x_\alpha})/2$ and $\text{spr}X_\alpha = (r_{x_\alpha} - l_{x_\alpha})/2$
with λ being the Lebesgue measure.

A particular case of distance:

$$D^2(A_\alpha, B_\alpha) = (1-\tau)(\text{mid}A_\alpha - \text{mid}B_\alpha)^2 + \tau(\text{spr}A_\alpha - \text{spr}B_\alpha)^2$$

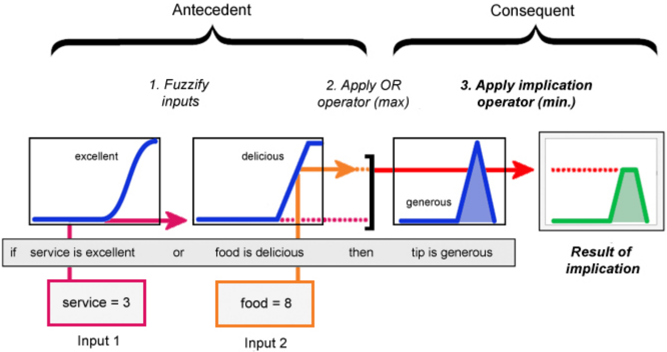
Fuzzy numbers can be used in the definition of **fuzzy variables**, i.e. a linguistic variable whose levels are represented as fuzzy numbers.

For instance, the variable *socio-economic status* (SES) can be represented in terms of a fuzzy variable:



Fuzzy variables are of most importance in **fuzzy systems**.

Fuzzy variables can be combined using fuzzy logic rules to make appropriate inference [38].



Source: `https://tinyurl.com/8s4cnbfz`

Fuzzy systems can be used in many practical applications, including:

- classification algorithms
- neural networks
- adaptive learning algorithms
- artificial intelligence
- numerical solvers
- image processing
- ...

FST across science

The applications of FST span across many scientific disciplines, including both theoretical and applied.

With regards to social and behavioral sciences, for instance, FST has successfully been used in:

- quantitative sociology [40]
- economics [3]
- education [32]
- social statistics [34]

In these cases, the role of FST has been to improve standard quantitative methods.

For instance, Lalla et al. (2005) [32] used a fuzzy set system to evaluate *teaching activity*. In particular, they set up a fuzzy system to compute the final composite teaching score by aggregating several Likert-type indicators.

Similarly, Betti & Verma (2008) [2] used FST to create a multidimensional *indicator of poverty* which resulted from the aggregation of empirical indicators about social and economic situations.

Likewise, Ragin (2000) [37] introduced the *fuzzy-Qualitative Comparative Analysis* (QCA), a qualitative technique to model causal pathways in a set of variables.

Outline

An elementary example

While probability formalizes the **randomness** of an aleatory experiment (i.e., uncertainty about the occurrence of random events), possibility formalizes the **epistemic imprecision** of a random experiment (i.e., uncertainty about the way random events are defined).

a proposition is uncertain if it involves a stochastic process; [. . .] an exact proposition may be uncertain ("it will be 4°C tomorrow"), and a proposition which is completely certain may be linguistically inexact ("it is warm now") [45]

Consider the experiment of tossing a (fair) coin with two possible outcomes: HEAD or TAIL. Over a sequence of independent tosses, the observer registered on a sheet whether the coin is HEAD (H) or TAIL (T).

$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	...
				T?	H?	H?	...

For certain tosses (i.e., $i \geq 5$) the observer got distracted and she/he was uncertain about the outcome (T or H ?).

Then, she/he decided to report the outcomes linguistically:

$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	...
				approx T	approx H	approx H	...

The coin tossing experiment can be usually modeled using a Bernoulli random variable $X \sim \mathcal{B}ern(x;\pi)$, with $\text{supp}(X) = \{0,1\}$ and

$$X = 0 : \text{coin is } \textsc{head} \qquad X = 1 : \text{coin is } \textsc{tail}$$

However, to deal with fuzzy outcomes (*was it head or tail?*), $\text{supp}(X)$ needs to be defined in terms of fuzzy sets

$$\text{supp}(X) = \{\xi_{\tilde{0}}, \xi_{\tilde{1}}\}$$

Abusing notation slightly, the original probability space

$$(\{\xi_{\tilde{0}}, \xi_{\tilde{1}}\}, \mathbb{P}_X)$$

enlarges to cope with the new source of uncertainty (fuzzy probability space).

An elementary example

Tossing a coin

$$\Big(\overbrace{\{\xi_{\tilde{0}}, \xi_{\tilde{1}}\}}^{\text{fuzziness}}, \underbrace{\mathbb{P}_X}_{\text{randomness}}\Big)$$

In this case, the observer is **uncertain** about what

- the events 0 (H) and 1 (T) refer to (*fuzziness*)
- the occurrence of such events over a repeated experiment (*randomness*)

An elementary example

Tossing a coin

Let

$$\xi_0(x) = \begin{cases} 1.0 & x = 0 \\ 0.0 & x = 1 \end{cases} \qquad \xi_1(x) = \begin{cases} 0.0 & x = 0 \\ 1.0 & x = 1 \end{cases}$$

be the *crisp sets* for the events $X = 0$ ($\textsc{h}$) and $X = 1$ ($\textsc{t}$).

$$\xi_{\tilde{0}}(x) = \begin{cases} 0.9 & x = 0 \\ 0.2 & x = 1 \end{cases} \qquad \xi_{\tilde{1}}(x) = \begin{cases} 0.1 & x = 0 \\ 0.8 & x = 1 \end{cases}$$

be the *fuzzy sets* for the events $X \stackrel{\sim}{=} 0$ (approx $\textsc{h}$) and $X \stackrel{\sim}{=} 1$ (approx $\textsc{t}$).

An elementary example

Tossing a coin

In the Bernoulli case, $\mathbb{P}_X(x) = \pi^x(1-\pi)^{1-x}$ with $\pi \in (0,1)$ as usual.

The probability of an event is equal to (e.g., see [4]):

$$\mathbb{P}_X(X = x) = \sum_{x \in \{0,1\}} \xi_{\widetilde{x}}(x)\mathbb{P}_X(x)$$

where the Bernoulli probabilities are weighted by the membership function of the fuzzy sets.

An elementary example

Tossing a coin

$i = 1$	$i = 2$	$i = 3$	...
			...

With $\pi = 0.5$,

the probability of HEAD is

$$\begin{aligned}\mathbb{P}_X(X = 0) &= 1 \cdot (0.5^0 \cdot (1 - 0.5)^1) + 0 \cdot (0.5^1 \cdot (1 - 0.5)^0) \\ &= 0.50\end{aligned}$$

whereas the probability of TAIL is

$$\begin{aligned}\mathbb{P}_X(X = 1) &= 0 \cdot (0.5^0 \cdot (1 - 0.5)^1) + 1 \cdot (0.5^1 \cdot (1 - 0.5)^0) \\ &= 0.50\end{aligned}$$

An elementary example

Tossing a coin

...	$i = 5$	$i = 6$	$i = 7$	...
...	approx T	approx H	approx H	...

With $\pi = 0.5$,

the probability of "approximately HEAD" is

$$\begin{aligned}\mathbb{P}_X(X \overset{\sim}{=} 0) &= 0.9 \cdot (0.5^0 \cdot (1-0.5)^1) + 0.2 \cdot (0.5^1 \cdot (1-0.5)^0) \\ &= 0.9 \cdot 0.5 + 0.2 \cdot 0.5 = 0.55\end{aligned}$$

whereas the probability of "approximately TAIL" is

$$\begin{aligned}\mathbb{P}_X(X \overset{\sim}{=} 1) &= 0.1 \cdot (0.5^0 \cdot (1-0.5)^1) + 0.8 \cdot (0.5^1 \cdot (1-0.5)^0) \\ &= 0.1 \cdot 0.5 + 0.8 \cdot 0.5 = 0.45\end{aligned}$$

Modeling fuzziness as source of uncertainty, requires introducing the notion of **monotone measures** [30] which allows for representing fuzziness in terms of **possibility measure**.

Broadly speaking, fuzzy membership function $\xi_{\widetilde{X}}$ can be interpreted as possibility measure [13], which in turn belongs to the family of **imprecise probabilities** (in particular, *p-boxes*).

For further details, refer to [1].

Since Zadeh's definition of probability of a fuzzy event [44], there have been a plethora of attempts to combine fuzziness and randomness (e.g., see [4, 45, 21, 39, 1, 9, 28, 7]).

The findings provided by the *SMIRE group* (e.g., M. Gil, I. Couso, P. Teran) are the most prominent in this debate.

Further details: `http://bellman.ciencias.uniovi.es/smire/`

A **fuzzy random variable** is a random variable whose possible outcomes are fuzzy numbers instead of real numbers.

(Didactically speaking) there are two main approaches (although things are more complicated in reality):

- **Ontic**: The focus is on the non-standard random mechanism that produces fuzzy outcomes. In this case, the phenomenon being modeled is fuzzy *ontologically* (i.e., fuzzy in nature).

A **fuzzy random variable** is a random variable whose possible outcomes are fuzzy numbers instead of real numbers.

(Didactically speaking) there are two main approaches (although things are more complicated in reality):

- **Ontic**: The focus is on the non-standard random mechanism that produces fuzzy outcomes. In this case, the phenomenon being modeled is fuzzy *ontologically* (i.e., fuzzy in nature).

- **Epistemic**: Fuzzy numbers represent descriptions of unobserved (latent) underlying crisp random variables. In this case, the phenomenon being modeled is fuzzy because of a lack of ability/knowledge in observing the true but latent outcomes (i.e., fuzzy in knowledge). See slides: 34-35.

A **fuzzy random variable** is a random variable whose possible outcomes are fuzzy numbers instead of real numbers.

(Didactically speaking) there are two main approaches (although things are more complicated in reality):

- **Ontic**: The focus is on the non-standard random mechanism that produces fuzzy outcomes. In this case, the phenomenon being modeled is fuzzy *ontologically* (i.e., fuzzy in nature).

- **Epistemic**: Fuzzy numbers represent descriptions of unobserved (latent) underlying crisp random variables. In this case, the phenomenon being modeled is fuzzy because of a lack of ability/knowledge in observing the true but latent outcomes (i.e., fuzzy in knowledge). See slides: 34-35.

  $\rightarrow$ Related to **censored** and **coarse data** [6, 35]

Let the family of real fuzzy numbers be

$$\mathcal{F}(\mathbb{R}) = \{\xi_{\widetilde{A}} : \mathbb{R} \to [0,1] \mid \xi_{\widetilde{A}_\alpha} \in \mathcal{K}(\mathbb{R}), \alpha \in [0,1]\}$$

with $\mathcal{K}(\mathbb{R})$ being the family of all non-empty compact intervals.

Consider two fuzzy numbers $\tilde{A}, \tilde{B} \in \mathcal{F}(\mathbb{R})$, then (see slides 23-26):

- the α cut-based operations of summation $\tilde{A} + \tilde{B}$ (Minkowski'sum) and (scalar) product $b\tilde{A}$ are applied element-wise on the elements of intervals
- the difference $\tilde{A} -_H \tilde{B}$ is the Hukuhara difference
- $D_\tau^\lambda(\tilde{A}, \tilde{B})$ is used to measure distances

A fuzzy random variable is a Borel measurable mapping $\tilde{X}$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the metric space $(\mathcal{F}(\mathbb{R}), D_\tau^\lambda)$.
Note: given the α-cuts representation, $\tilde{X}$ works level-wise.

Given a collection of fuzzy rvs $\mathcal{X} = (\tilde{X}_1, \ldots, \tilde{X}_n)$:

- $\tilde{\mu} = \mathbb{E}\left[\mathcal{X}\right] = \Big[\mathbb{E}\left[\mathrm{mid}\mathcal{X}_\alpha\right] - \mathbb{E}\left[\mathrm{spr}\mathcal{X}_\alpha\right], \mathbb{E}\left[\mathrm{mid}\mathcal{X}_\alpha\right] + \mathbb{E}\left[\mathrm{spr}\mathcal{X}_\alpha\right]\Big] \in \mathcal{F}(\mathbb{R})$
- $\sigma^2_{\mathcal{X}} = \mathbb{V}\mathrm{ar}\left[\mathcal{X}\right] = \mathbb{E}\left[D_\tau^\lambda(\mathcal{X}, \tilde{\mu})^2\right] \in \mathbb{R}^+$
- If $\mathcal{Y}$ is also available:

  $\sigma_{\mathcal{X},\mathcal{Y}} = \mathbb{C}\mathrm{ov}\left[\mathcal{X}, \mathcal{Y}\right]_{|D_\tau^\lambda} = \mathbb{C}\mathrm{ov}\left[\mathrm{mid}\mathcal{X}, \mathrm{mid}\mathcal{Y}\right] + \mathbb{C}\mathrm{ov}\left[\mathrm{spr}\mathcal{X}, \mathrm{spr}\mathcal{Y}\right] \in \mathbb{R}$

Note:

- $\tilde{\mu}$ preserves all the main properties from the crisp case (e.g., additivity, equivariance under translation and product by a scalar)
- $\sigma^2_{\mathcal{X}}$ preserves all the main properties from the crisp case (e.g., it vanishes with degenerate distribution, invariance under translation, additivity under independence)
- $\sigma_{\mathcal{X},\mathcal{Y}}$ preserves some properties from the crisp case (e.g., it vanishes for independent random variables)

Some limitations of the Ontic approach if compared to standard statistics:

- Use of Hukuhara operator to approximate difference between fuzzy numbers
- Lack of a general total ranking between fuzzy numbers
- No flexible models for fuzzy random variables
- Limit theorems do not always apply for fuzzy random variables (lack of easy-to-apply statistical inference)

Outline

Overview

To generalize what we have said in the previous section, regardless of both epistemic and ontic approaches to statistics, we can recognize that fuzzy numbers allow for representing a systematic and non-random uncertainty associated with *data*, *parameters*, or statistical *hypotheses*.

Overview

Consider a generic (parametric) statistical model of the form

$$\mathcal{M} = \{F(\mathbf{y};\boldsymbol{\theta}), \boxed{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \subset \mathbb{R}^p, \boxed{\mathbf{y} \in \mathcal{Y}^n}\}$$

Fuzzy numbers can be introduced in this context to:

- represent the **parameter space** $\boldsymbol{\Theta}$
- represent the **sample space** $\mathcal{Y}^n$

In both cases, the statistical model needs to deal with two different sources of uncertainty at least: the **aleatoric uncertainty** and the **fuzziness** (provided by fuzzy numbers).

When the parameter space is a subset of fuzzy numbers, i.e. $\boldsymbol{\Theta} \subset \mathcal{F}(\mathbb{R})^p$, we may have several statistical models where **parameters are fuzzy sets**:

- Fuzzy clustering methods (e.g., [12])
- Fuzzy regression (e.g., [8])
- Fuzzy time series models (e.g., [8])
- ...

When the sample space is a subset of fuzzy numbers, e.g. $\mathcal{Y}^n \subset \mathcal{F}(\mathbb{R})^n$ or $\mathcal{Y}^n \subset \mathcal{F}(\mathbb{N})^n$, we may have several statistical models where **data are fuzzy sets**:

- Fuzzy linear models (e.g., [5, 8])
- Fuzzy time series models (e.g., [8])
- Fuzzy component analysis (e.g., [23])
- ...

In these cases, sufficient statistics for estimating model parameters are based on fuzzy data as well. The inferential mechanism has to be generalized according to this new type of data.

Overview

Fuzzy hypothesis testing

Statistical hypothesis testing can also be performed using a set of *fuzzy hypotheses* over $\boldsymbol{\Theta}$:

- Fuzzy Neyman-Pearson lemma for Fuzzy Most Powerful Test (fuzzy-MP) and Fuzzy Uniformly Most Powerful Test (fuzzy-UMP) [41, 43]
- Fuzzy p-values [17, 20]
- Minimax fuzzy tests [36]
- Fuzzy Bayesian inference [18, 20]

Example

Let $(X_1, \ldots, X_i, \ldots, X_n)$ be a random sample with $X_i \sim \mathcal{N}(x; \theta, \sigma^2 = \sigma_0^2)$ where θ is the parameter (i.e., the mean) to be tested under the following *fuzzy hypothesis system*:

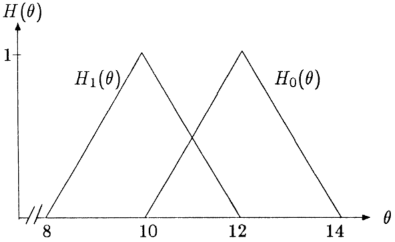
$$\begin{cases} \mathcal{H}_0 : \theta \overset{\sim}{=} 12 \\ \mathcal{H}_1 : \theta \overset{\sim}{=} 10 \end{cases}$$

where $\overset{\sim}{=}$ means *approximately* (in a fuzzy sense).

The hypotheses over $\boldsymbol{\Theta}$ can be expressed using triangular fuzzy numbers.

Example

$$\begin{cases} \mathcal{H}_0 : \theta \cong 12 \\ \mathcal{H}_1 : \theta \cong 10 \end{cases}$$



Source: [41]

Example

In this case, error probabilities, power and test functions can be computed by including fuzzy numbers into the likelihood ratio:

$$\lambda(x) = \frac{\int N(x;\theta,\sigma_0^2)\xi_{\widetilde{H}_0}(\theta)}{\int N(x;\theta,\sigma_0^2)\xi_{\widetilde{H}_1}(\theta)} \leq k$$

Further details: [41]

Fuzzy k-means

Fuzzy clustering has been one of the first data analysis technique which has directly used fuzzy sets into its mechanisms. Since the first attempt by Dunn [14], a number of improvements and generalizations have been proposed over the years [16, 29].

Fuzzy k-means

In general, let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a (crisp) data matrix and let $g = 1, \ldots, G$ be the number of (unobserved) clusters. Then, the **fuzzy k-means algorithm** seeks for the best fuzzy partition of n units into G groups by solving the following problem:

$$\underset{\boldsymbol{\Xi},\mathbf{H}}{\text{minimize}} \quad \sum_i \sum_g \xi_{ig}^{\lambda} \left\| \mathbf{x}_i - \mathbf{h}_g \right\|_2^2$$

$$\text{subject to:} \quad (1)\ \boldsymbol{\xi}_i \in [0,1]^g \quad (2)\ \boldsymbol{\xi}_i \mathbf{1}_G = 1$$

where $\boldsymbol{\Xi}_{n \times G}$ is the matrix of fuzzy degrees of membership of the i-th unit to the g-th cluster whereas $\mathbf{H}_{G \times p}$ is the matrix of **centroids**. Note that λ is the tuning parameter of the algorithm (usually, $\lambda \in [1.5, 2]$).

Fuzzy k-means

The basic formulation has been extended so as to:

□ generalize beyond the standard spherical cluster solution (i.e., *Gustafson and Kessel*'s solution)

$$\underset{\boldsymbol{\Xi},\mathbf{H},\mathbf{V}_1,\ldots,\mathbf{V}_G}{\text{minimize}} \quad \sum_i \sum_g \xi_{ig}^{\lambda}(\mathbf{x}_i - \mathbf{h}_g)^T \mathbf{V}_g^{-1}(\mathbf{x}_i - \mathbf{h}_i)$$

$$\text{subject to:} \quad (1)\ \boldsymbol{\xi}_i \in [0,1]^g \quad (2)\ \boldsymbol{\xi}_i \mathbf{1}_G = 1 \quad (3)\ |\mathbf{V}_g| = \rho_g$$

Fuzzy k-means

The basic formulation has been extended so as to:

☐ remove the ambiguous tuning parameter λ (i.e., Li and *Mukaidono*'s solution)

$$\underset{\boldsymbol{\Xi},\mathbf{H}}{\text{minimize}} \quad \sum_i \sum_g \xi_{ig} \left\|\mathbf{x}_i - \mathbf{h}_g\right\|_2^2 + \pi \sum_i \sum_g \xi_{ig} \log \xi_{ig}$$

$$\text{subject to:} \quad {\scriptstyle (1)}\ \boldsymbol{\xi}_i \in [0,1]^g \quad {\scriptstyle (2)}\ \boldsymbol{\xi}_i \mathbf{1}_G = 1$$

with $\pi > 0$ being the *temperature* of the algorithm (it weights the entropy part of the objective function).

Fuzzy k-medoids

The matrix of centroids $\mathbf{H}_{G\times p}$ can be replaced with the matrix of **medoids** $\mathbf{J}_{G\times p}$:

$$\underset{\boldsymbol{\Xi},\mathbf{J}}{\text{minimize}} \quad \sum_i \sum_g \xi_{ig}^{\lambda} \left\| \mathbf{x}_i - \mathbf{j}_g \right\|_2^2$$

$$\text{subject to:} \quad {\scriptstyle(1)}\ \boldsymbol{\xi}_i \in [0,1]^g \quad {\scriptstyle(2)}\ \boldsymbol{\xi}_i \mathbf{1}_G = \mathbf{1}_n$$

Note: given a set $S = \{s_1, \ldots, s_n\}$ with a distance (or dissimilarity) function $d()$, the medoid is the point

$$\tilde{x} = \arg\min_{q\in S}\ d(y, s_i)$$

As for the non-fuzzy version of clustering techniques, fuzzy solutions of k-means and k-medoids can be assessed by means of indices or graphical analyses.

- **modified Partition index**

$$1 - \frac{G}{G-n} 1 - \sum_i \sum_g \frac{\xi_{ig}^2}{n} \quad \text{(it needs to be maximized over } g\text{)}$$

- **Partition entropy**

$$-\sum_i \sum_g \frac{1}{n} \xi_{ig} \log \xi_{ig} \quad \text{(it needs to be minimized over } g\text{)}$$

As for the non-fuzzy version of clustering techniques, fuzzy solutions of k-means and k-medoids can be assessed by means of indices or graphical analyses.

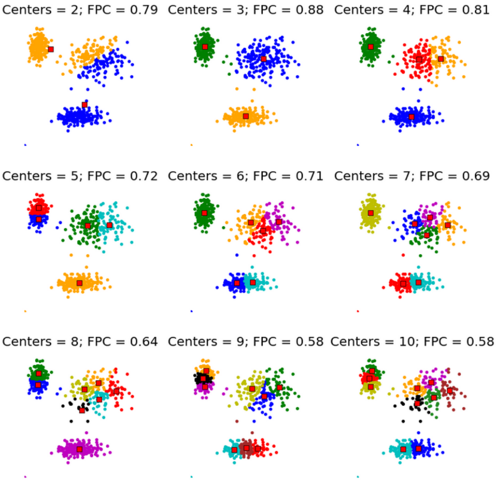
- **fuzzy Silhouette index**

$$\frac{\sum_i (\xi_{ig_1} - \xi_{ig_2})\ \mathsf{sil}(g)}{\sum_i (\xi_{ig_1} - \xi_{ig_2})} \quad \text{(it needs to be maximized over } g\text{)}$$

where sil(g) is the standard Silhouette index whereas g_1 and g_2 are the first and the second largest elements of the i-th row of $\boldsymbol{\Xi}$.

Fuzzy k-means

Example



Source: `https://tinyurl.com/32y5t2pj`

Fuzzy regression

A second technique, as old as fuzzy clustering, is that of **fuzzy regression**. In the literature, there have been thousands of proposals [8], all of them revolving around the following generalizations:

(a) $\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$ (Crisp response, crisp predictors, fuzzy coefficients)

(b) $\tilde{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (Fuzzy response, crisp predictors, crisp coefficients)

(c) $\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (Fuzzy response, fuzzy predictors, crisp coefficients)

(d) $\tilde{\mathbf{y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$ (Fuzzy response, fuzzy predictors, fuzzy coefficients)

Regression problems have been solved under least squares, non-linear programming, and evolutionary methods.

The simplest formulation for $\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$ has been provided by Tanaka [42]:

$$\begin{aligned} y_i &= \tilde{\beta}_0 + \sum_j x_{ij}\tilde{\beta}_j + \epsilon_i \\ &= (m^{\beta_0}, s^{\beta_0}) + \sum_j x_{ij}(m_j^{\beta_j}, s_j^{\beta_j}) + \epsilon_i \end{aligned}$$

where the regression coefficients have been expressed as *symmetric triangular fuzzy numbers*, i.e. $\tilde{\beta} = (m, s)_{LR}$.

Note that the linear model predicts outcomes in terms of fuzzy numbers, i.e. $\hat{\mathbf{y}} = \hat{\tilde{\mathbf{y}}}$.

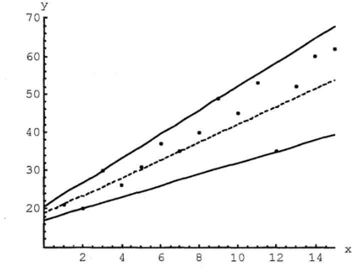
The regression model is determined by minimizing the *sum of spreads* of the estimated fuzzy outputs:

$$\min_{\mathbf{m},\mathbf{s}} \sum_i \sum_j |x_{ij}| c_j$$

$$\text{subject to:} \quad {\scriptstyle (1)}\ s_j \geq 0 \quad (2)\ y_i \in \mathcal{S}(\hat{\tilde{y}}_i)_h \quad i = 1, \ldots, n$$

where cons. (2) indicates that the predicted outcomes $\hat{\tilde{\mathbf{y}}}$ have to lie inside the h-level sets of the estimated regression lines.

No.(j)	1	2	3	4	5	6	7	8
x	1	2	3	4	5	6	7	8
y	21	20	30	26	31	37	35	40
No.(j)	9	10	11	12	13	14	15	
x	9	10	11	12	13	14	15	
y	49	45	53	35	52	60	62	



Source: [33]

Generalization of the Tanaka's proposal have been made so as to include asymmetric fuzzy numbers, fuzzy outcome and predictors. For further details, see: [33].

Consider a sample of fuzzy triangular numbers $\tilde{\mathbf{y}} = ((m_1, l_1, r_1), \ldots, (m_n, l_n, r_n))_{LR}$ and a matrix of crisp predictors $\mathbf{X}_{n \times p}$. Then, given the LR parametric representation, the following (*non-interactive*) linear model can be formulated [15]:

$$
\begin{aligned}
\mathbf{m} &= \mathbf{X}\boldsymbol{\beta}_m + \boldsymbol{\epsilon}_m \\
\mathbf{l} &= \mathbf{X}\boldsymbol{\beta}_l + \boldsymbol{\epsilon}_l \\
\mathbf{r} &= \mathbf{X}\boldsymbol{\beta}_r + \boldsymbol{\epsilon}_r
\end{aligned}
$$

Alternatively, a (*interactive*) linear model could also be developed:

$$\mathbf{m} = \overbrace{\mathbf{X}\boldsymbol{\beta}_m}^{\mathbf{m}^*} + \boldsymbol{\epsilon}_m$$
$$\mathbf{l} = \mathbf{m}^* \beta_l + \boldsymbol{\epsilon}_l$$
$$\mathbf{r} = \mathbf{m}^* \beta_r + \boldsymbol{\epsilon}_r$$

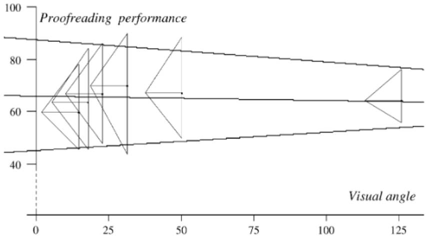
In both cases, the regression problem is solved by formulating the following (unconstrained) problem:

$$\min_{\boldsymbol{\beta}} \quad \|\mathbf{m} - \mathbf{m}^*\|_2^2 + \|\mathbf{l} - \mathbf{l}^*\|_2^2 + \|\mathbf{r} - \mathbf{r}^*\|_2^2$$

which can be constrained by letting the spread components to be non-negative:

$$\mathbf{l}^* \geq \mathbf{0}_n \qquad \text{and} \qquad \mathbf{r}^* \geq \mathbf{0}_n$$

Dataset		Parameter estimates
X	$Y \equiv (c, p, q)$	
14.8	(59.7, 13.7, 18.3)	
18.0	(63.5, 17.5, 20.5)	$\hat{\mathbf{a}} = (65.940128, -0.016853)'$
22.9	(66.8, 18.8, 19.2)	$\hat{b} = 5.1076942$
31.5	(70, 26, 20)	$\hat{d} = -316.155$
50.3	(67, 17, 21)	$\hat{g} = 3.9716534$
126.0	(64.2, 8.2, 11.8)	$\hat{h} = -240.4851$



Source: [15]

Generalization of the least squares fuzzy regression have been made so as to include trapezoidal fuzzy numbers, crisp outcome and fuzzy predictors.
For further details, see: [15].

Consider the simplest case of the Normal linear model:

$$\mathcal{M} = \left\{\mathcal{N}(\mathbf{y};\boldsymbol{\mu},\mathbf{I}\sigma^2), \boldsymbol{\theta} = \{\boldsymbol{\mu},\sigma^2\} \subset \mathbb{R}^n \times \mathbb{R}^+, \mathbf{y} \in \mathbb{R}^n\right\}$$

where the linear predictor (mean) of the model is a linear combination of p predictors $\mathbf{X}_{n\times p}$:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

As usual, the interest lies in identifying the true model $\mathcal{M}^0$ which has generated the sample data $\mathbf{y}$ (i.e., estimate the unknown parameters $\{\boldsymbol{\beta},\sigma^2\}$).

Now, assuming that $\mathbf{y}$ cannot be observed directly due to non-random sources of uncertainty (*post-sampling errors*) but the fuzzy sample $\tilde{\mathbf{y}}$ is instead available.

Note that the interest still lies in identifying the true (non-fuzzy) model $\mathcal{M}^0$.

The parameters $\{\boldsymbol{\beta}, \sigma^2\}$ of the Normal linear model with fuzzy observations can be estimated, for instance, using *maximum likelihood* adapted to deal with fuzzy numbers [22].

The **fuzzy likelihood function** is as follows:

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2; \tilde{\mathbf{y}}) = \prod_{i=1}^{n} \int_{\text{supp}(\tilde{y}_i)} \xi_{\widetilde{y_i}}(y) \mathcal{N}(y; \boldsymbol{\beta}, \sigma^2) \; dy$$

where $(\xi_{\widetilde{y_1}}, \ldots, \xi_{\widetilde{y_i}}, \ldots, \xi_{\widetilde{y_n}})$ is the sample of fuzzy data (e.g., trapezoidal fuzzy numbers).

Fuzzy regression

Maximum Likelihood regression with fuzzy data

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2; \tilde{\mathbf{y}}) = \prod_{i=1}^{n} \int_{\mathrm{supp}(\tilde{y}_i)} \xi_{\widetilde{y}_i}(y) \mathcal{N}(y; \boldsymbol{\beta}, \sigma^2)\ dy$$

fuzziness of data through fuzzy numbers

Fuzzy regression

Maximum Likelihood regression with fuzzy data

$$\mathcal{L}(\boldsymbol{\beta},\sigma^2;\tilde{\mathbf{y}}) = \prod_{i=1}^{n} \int_{\text{supp}(\tilde{y}_i)} \xi_{\tilde{y}_i}(y)\mathcal{N}(y;\boldsymbol{\beta},\sigma^2)\ dy$$

fuzziness of data through fuzzy numbers

randomness of data through the probabilistic model

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

Given a candidate $\boldsymbol{\theta}'$, the EM algorithm iterates between:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right]$
- **M-step**
  Maximize $\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y})$ by replacing $\mathbf{y}$ with $\mathbf{y}^*$

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right]$

  $\rightarrow$ Likelihood of the Normal linear model with no fuzzy observations

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right]$

  $\rightarrow$ Likelihood of the Normal linear model with no fuzzy observations

  $\rightarrow$ filtered data: what we would expect to observe if fuzziness was not present in the data

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right] =$

$$= \int y \; \frac{\xi_{\widetilde{y}_i}(y)\mathcal{N}(y; \boldsymbol{\theta}')}{\int \xi_{\widetilde{y}_i}(z)\mathcal{N}(y; \boldsymbol{\theta}') \; dz} \; dy$$

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right] =$

$$= \int y \; \frac{\xi_{\widetilde{y}_i}(y)\mathcal{N}(y; \boldsymbol{\theta}')}{\int \xi_{\widetilde{y}_i}(z)\mathcal{N}(y; \boldsymbol{\theta}') \; dz} \; dy$$

$\rightarrow$ Normal density function *conditioned* on fuzzy numbers!

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right] =$

$$= \int y \, \frac{\xi_{\widetilde{y}_i}(y)\mathcal{N}(y; \boldsymbol{\theta}')}{\underbrace{\int \xi_{\widetilde{y}_i}(z)\mathcal{N}(y; \boldsymbol{\theta}') \; dz}_{\text{normalization constant}}} \, dy$$

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **E-step**
  Compute $\mathbf{y}^* = \mathbb{E}_{\boldsymbol{\theta}'}\left[\ln \mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) | \tilde{\mathbf{y}}\right] =$

$$= \int y \, \frac{\overbrace{\xi_{\widetilde{y}_i}(y)}^{\text{fuzziness}} \, \overbrace{\mathcal{N}(y; \boldsymbol{\theta}')}^{\text{randomness}}}{\underbrace{\int \xi_{\widetilde{y}_i}(z) \mathcal{N}(y; \boldsymbol{\theta}') \, dz}_{\text{normalization constant}}} \, dy$$

The parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2\}$ can be estimated via Expectation Maximization (EM) [10]:

- **M-step**
  Maximize ln $\mathcal{L}(\boldsymbol{\beta}, \sigma^2; \mathbf{y})$ by replacing $\mathbf{y}$ with $\mathbf{y}^*$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}^*$$

$$\hat{\sigma}^2 = \frac{1}{n}\left(\mathbf{y}^* - \mathbf{X}\hat{\beta}\right)^T\left(\mathbf{y}^* - \mathbf{X}\hat{\beta}\right)$$

  where $\mathbf{y}^*$ are the observations filtered from fuzziness.

Fuzzy bootstrap
Overview

Fuzzy random variables constitute a well-founded approach to deal with fuzziness in statistical data analysis. However, due to the limitations on asymptotic results useful for doing inference (see slide 39), inference with fuzzy rvs is still an open issue.

To overcome some of the current limitations, bootstrap techniques have been widely adopted in fuzzy statistics (e.g., [24]).

Two recent proposals will be briefly considered here:

- Bootstrap for *epistemic* fuzzy data [26]
- Bootstrap for *ontic* fuzzy data [25]

Suppose a sample $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$ of fuzzy observations (e.g., trapezoidal/triangular) is available given a collection of (latent/unobserved) crisp random variables $(X_1, \ldots, X_n)$. In this context, $\tilde{x}_i \in \mathcal{F}(\mathbb{R})$ and $[\tilde{x}_i]_\alpha$ denotes the α-cut of $\tilde{x}_i$ (bounded interval).

Then, for $b = 1, \ldots, B$ [25]:

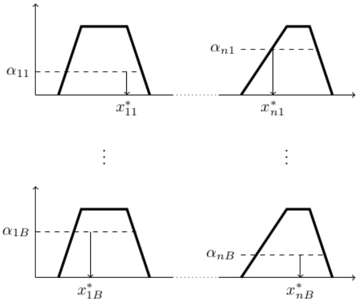
s_1 : $\alpha \sim \mathcal{U}(; 0, 1)$

s_2 : $\hat{x}_i^{(b)} \sim \mathcal{U}(; \min\ [\tilde{x}_i]_\alpha, \max\ [\tilde{x}_i]_\alpha)$

Do s_1-s_2 for $i = 1, \ldots, n$

The bootstrap sample $(\hat{x}_1, \ldots, \hat{x}_n)^{(b)}$ constitutes a random sample from $(X_1, \ldots, X_n)$ which can be used to compute statistics of interest $T_B(\hat{\mathbf{x}}^{(b)})$ as usual.

Graphical representation of steps s_1 and s_2 for trapezoidal fuzzy numbers:



Source: [25]

Fuzzy bootstrap

Epistemic approach

To reduce the variance of the bootstrap-based estimates, well-known Monte Carlo techniques (e.g., the *antithetic variates method*) can be generalized as well. For further details, see [25].

Fuzzy bootstrap

Ontic approach

Let $\mathcal{X} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ be a collection of fuzzy rvs with $\tilde{\mathbf{x}}$ being a sample of n trapezoidal fuzzy numbers.

The i-th observation is parameterized as $\tilde{x}_i = (m, s, l, r)$, with m and s denoting the center and the width of the core, l and r denoting the usual left and right spreads.

The following quantities can be used to synthesize $\tilde{x}_i$ (*canonical representation*) [26]:

- $\text{Val}(\tilde{x}) = c + (r - l)\frac{1}{6}$ (Location of the fuzzy number)
- $\text{Amb}_L(\tilde{x}) = \frac{1}{2}s + \frac{1}{6}l$ (Left ambiguity of the fuzzy number)
- $\text{Amb}_U(\tilde{x}) = \frac{1}{2}s + \frac{1}{6}r$ (Right ambiguity of the fuzzy number)
- $\text{EV}(\tilde{x}) = c + (r - l)\frac{1}{4}$ (Exp. value of the fuzzy number)
- $\text{w}(\tilde{x}) = s + (r + l)\frac{1}{4}$ (Width of the fuzzy number)

In this case, the bootstrap technique can be applied on the canonical representation of the fuzzy sample $\tilde{\mathbf{x}}$. Note that, although the fuzzy observations do not need to obey a particular shape, the bootstrap samples are always represented in terms of trapezoidal fuzzy numbers.

Several algorithms can be defined based on preserving some properties of the canonical representation, e.g. **VA**, **VAA**, **VAF**.

Fuzzy bootstrap

Ontic approach

For the sake of simplicity, the **VAA algorithm** is reproduced here.

Require: Fuzzy sample $x_1, \dots, x_n \in \mathbb{F}(\mathbb{R})$
Ensure: A flexible bootstrap sample

for $i = 1$ to n **do**
  Compute $\mathrm{Val}(x_i), \mathrm{Amb}_L(x_i), \mathrm{Amb}_U(x_i)$
end for
for $i = 1$ to n **do**
  Generate randomly (with equal probabilities) a triple $(\mathrm{Val}^*, \mathrm{Amb}_L^*, \mathrm{Amb}_U^*)$ from

$$\left\{ \left(\mathrm{Val}(x_1), \mathrm{Amb}_L(x_1), \mathrm{Amb}_U(x_1)\right), \dots, \left(\mathrm{Val}(x_n), \mathrm{Amb}_L(x_n), \mathrm{Amb}_U(x_n)\right) \right\}$$

  $c^* \leftarrow \mathrm{Val}^* + \mathrm{Amb}_U^* - \mathrm{Amb}_L^*$
  Generate s^* from the uniform distribution on the interval

$$\left[0, 2 \min \left\{ \mathrm{Amb}_L^*, \mathrm{Amb}_U^* \right\}\right]$$

  $l^* \leftarrow 6\mathrm{Amb}_L^* - 3s^*$
  $r^* \leftarrow 6\mathrm{Amb}_U^* - 3s^*$
  $x_i^* \leftarrow x_i^*(c^*, s^*; l^*, r^*)$
end for

Source: [26]

Outline

Concluding remarks

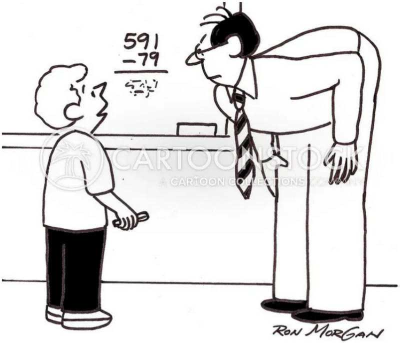
- Fuzzy set theory is a broad mathematical theory dealing with phenomena affected by uncertainty (i.e., ambiguity or vagueness)

- Fuzzy numbers and their generalizations (e.g., type-2 fuzzy numbers, fuzzy quaternions) introduce more flexibility in many mathematical problems dealing with this type of uncertainty

Concluding remarks

- When coupled with standard probability theory, fuzzy numbers allow for generalizing statistical models as well as statistical inference to cope with different sources of uncertainty in the same time

- Up to now, a common and unified formal representation subsuming all the findings from fuzzy statistics, fuzzy probability theory, and fuzzy data analysis is still missing



"It's fuzzy math."

antonio.calcagni@unipd.it

https://unipd.link/acalcagni

[1] Augustin, T., Coolen, F. P., De Cooman, G., and Troffaes, M. C.
Introduction to imprecise probabilities.
John Wiley & Sons, 2014.

[2] Betti, G., and Verma, V.
Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective.
Statistical Methods and Applications 17, 2 (2008), 225–250.

[3] Buckley, J. J., Eslami, E., and Feuring, T.
Fuzzy mathematics in economics and engineering, vol. 91.
Physica, 2013.

[4] Buoncristiani, J.
Probability on fuzzy sets and ϵ-fuzzy sets.
Journal of mathematical analysis and applications 96, 1 (1983), 24–41.

[5] Calcagnì, A., and Lombardi, L.
Modeling random and non-random decision uncertainty in ratings data: a fuzzy beta model.
AStA Advances in Statistical Analysis (2021), 1–29.

[6] Cattaneo, M., and Wiencierz, A.
Robust regression with coarse data.

[7] Cattaneo, M. E.
The likelihood interpretation as the foundation of fuzzy set theory.
International Journal of Approximate Reasoning 90 (2017), 333–340.

[8] Chukhrova, N., and Johannssen, A.
Fuzzy regression analysis: systematic review and bibliography.
Applied Soft Computing 84 (2019), 105708.

[9] Couso, I., and Dubois, D.
Statistical reasoning with set-valued information: Ontic vs. epistemic views.
International Journal of Approximate Reasoning 55, 7 (2014), 1502–1518.

[10] Denœux, T.
Maximum likelihood estimation from fuzzy data using the em algorithm.
Fuzzy sets and systems 183, 1 (2011), 72–91.

[11] Dijkman, J., Van Haeringen, H., and De Lange, S.
Fuzzy numbers.
Journal of Mathematical Analysis and Applications 92, 2 (1983), 301–341.

[12] Döring, C., Lesot, M.-J., and Kruse, R.
Data analysis with fuzzy clustering methods.
Computational Statistics & Data Analysis 51, 1 (2006), 192–214.

[13] Dubois, D., and Prade, H.
Fundamentals of fuzzy sets, vol. 7.
Springer Science & Business Media, 2012.

[14] Dunn, J. C.
A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters.

[15] D'Urso, P.
Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data.
Computational Statistics & Data Analysis 42, 1-2 (2003), 47–72.

[16] Ferraro, M. B., and Giordani, P.
A toolbox for fuzzy clustering using the r programming language.
Fuzzy Sets and Systems 279 (2015), 1–16.

[17] Filzmoser, P., and Viertl, R.
Testing hypotheses with fuzzy data: the fuzzy p-value.
Metrika 59, 1 (2004), 21–29.

[18] Frühwirth-Schnatter, S.
On fuzzy bayesian inference.
Fuzzy sets and systems 60, 1 (1993), 41–58.

[19] Galanis, G. N.
Differentiability on semilinear spaces.
Nonlinear Analysis: Theory, Methods & Applications 71, 10 (2009), 4732–4738.

[20] Geyer, C. J., and Meeden, G. D.
Fuzzy and randomized confidence intervals and p-values.
Statistical Science (2005), 358–366.

[21] Gil, M. A.
A note on the connection between fuzzy numbers and random intervals.
Statistics & probability letters 13, 4 (1992), 311–319.

[22] Gil, M. A.
Sufficiency and fuzziness in random experiments.
Annals of the Institute of Statistical Mathematics 44, 3 (1992), 451–462.

[23] Giordani, P., and Kiers, H. A.
Principal component analysis of symmetric fuzzy data.
Computational statistics & data analysis 45, 3 (2004), 519–548.

[24] González-Rodríguez, G., Montenegro, M., Colubi, A., and Gil, M. Á.
Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data.
Fuzzy Sets and Systems 157, 19 (2006), 2608–2613.

[25] Grzegorzewski, P., and Romaniuk, M.
Bootstrap methods for fuzzy data.
In *International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets* (2020), Springer, pp. 28–47.

[26] Grzegorzewski, P., and Romaniuk, M.
Bootstrap methods for epistemic fuzzy data.
International Journal of Applied Mathematics and Computer Science 32, 2 (2022).

[27] Hanss, M.
Applied fuzzy arithmetic.
Springer, 2005.

[28] Hesamian, G., and Shams, M.
A note on fuzzy probability of a fuzzy event.
International Journal of Intelligent Systems 32, 7 (2017), 676–685.

[29] Höppner, F., Klawonn, F., Kruse, R., and Runkler, T.
Fuzzy cluster analysis: methods for classification, data analysis and image recognition.
John Wiley & Sons, 1999.

[30] Klir, G. J.
Uncertainty and information: foundations of generalized information theory.
Kybernetes 35, 7/8 (2006), 1297–1299.

[31] Kosko, B., and Toms, M.
Fuzzy thinking: The new science of fuzzy logic, vol. 288.
Hyperion New York, 1993.

[32] Lalla, M., Facchinetti, G., and Mastroleo, G.
Ordinal scales and fuzzy set systems to measure agreement: an application to the evaluation of teaching activity.
Quality and Quantity 38, 5 (2005), 577–601.

[33] Lee, H., and Tanaka, H.
Fuzzy approximations with non-symmetric fuzzy parameters in fuzzy regression analysis.
Journal of the Operations Research Society of Japan 42, 1 (1999), 98–112.

[34] Lemmi, A. A., and i, G.
Fuzzy set approach to multidimensional poverty measurement, vol. 3.
Springer Science & Business Media, 2006.

[35] Nguyen, H. T., and Wu, B.
Fundamentals of statistics with fuzzy data, vol. 198.
Springer, 2006.

[36] Parchami, A., Taheri, S. M., Viertl, R., and Mashinchi, M.
Minimax test for fuzzy hypotheses.
Statistical Papers 59, 4 (2018), 1623–1648.

[37] Ragin, C. C.
Fuzzy-set social science.
University of Chicago Press, 2000.

[38] Ross, T. J.
Fuzzy logic with engineering applications.
John Wiley & Sons, 2005.

[39] Singpurwalla, N. D., and Booker, J. M.
Membership functions and probability measures of fuzzy sets.
Journal of the American statistical association 99, 467 (2004), 867–877.

[40] Smithson, M.
Fuzzy set analysis for behavioral and social sciences.
Springer Science & Business Media, 2012.

[41] Taheri, S. M., and Behboodian, J.
Neyman-pearson lemma for fuzzy hypotheses testing.
Metrika 1, 49 (1999), 3–17.

[42] Tanaka, H., Hayashi, I., and Watada, J.
Possibilistic linear regression analysis for fuzzy data.
European Journal of Operational Research 40, 3 (1989), 389–396.

[43] Torabi, H., Behboodian, J., and Taheri, S. M.
Neyman–pearson lemma for fuzzy hypotheses testing with vague data.
Metrika 64, 3 (2006), 289–304.

[44] Zadeh, L. A.
Probability measures of fuzzy events.
Journal of mathematical analysis and applications 23, 2 (1968), 421–427.

[45] Zwick, R., and Wallsten, T. S.
Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models.
International Journal of Man-Machine Studies 30, 1 (1989), 69–111.