Fuzzy Data and Information Systems

REINHARD VIERTL
Vienna University of Technology
Department of Statistics and Probability Theory
Wiedner Hauptstrasze 8, 1040 Vienna
AUSTRIA
r.viertl@tuwien.ac.at

Abstract: Real quantitative data and other valuable informations are often not precise numbers but more or less non-precise. This kind of uncertainty is also called fuzziness, and the related information is called fuzzy information. The best up-to-date description of this kind of data is by so-called ”fuzzy numbers”. In order to include such data in databases the databases have to be able to store fuzzy numbers in a suitable way.

Key–Words: data bases, fuzzy data, fuzzy numbers, fuzzy vectors, information systems

1 Introduction

Many real data cannot be adequately described by precise real numbers. For example environmental data, quality of life data, and all measurements of continuous one-dimensional quantities.

Realistic measurement results from continuous quantities are always not precise multiples of the measurement unit, but more or less non-precise (fuzzy). This imprecision is different from measurement errors, and for realistic quantitative analysis imprecision has to be modelled and also represented in data bases. This is possible using the concept of fuzzy numbers.

In case of multivariate data (for example locations in regions in the context of environmental data) such data are often fuzzy. This kind of data can be modelled by so-called fuzzy vectors.

2 Fuzzy numbers

For the quantitative description of fuzzy data a generalization of real numbers is necessary which also includes data in form of intervals.

In generalization of real numbers the following specialization of fuzzy subsets of the real line $\mathbb{R}$ is useful.

Definition 1 *A fuzzy real number $x^\star$ is defined by its so-called* characterizing function $\xi(\cdot)$, *where $\xi(\cdot)$ is a real function of one real variable x obeying the following:*

- $0 \leq \xi(x) \leq 1 \qquad \forall \quad x \in \mathbb{R}$
- the support of $\xi(\cdot)$ is a bounded subset of $\mathbb{R}$
- for all $\delta \in (0,1]$ the so-called δ-cut $C_\delta[x^\star]$ defined by $C_\delta[x^\star] := \{x \in \mathbb{R}\colon \xi(x) \geq \delta\}$ is non-empty and a finite union of compact intervals, i. e.

$$C_\delta[x^\star] = \bigcup_{j=1}^{m_\delta} [a_{\delta,j}, b_{\delta,j}],$$

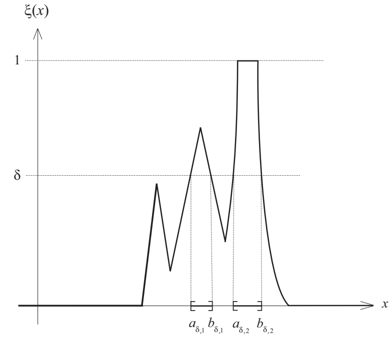
where $C_\delta[x^\star]$ can be degenerated to a single point, i. e. $[a,a]$.

This definition is more general than the usual one, but it is necessary in order to describe real data. For example fuzzy data obtained from color intensity pictures with monotone color intensity transition.

In figure 1 an example of the characterizing function of a fuzzy number is depicted.

Remark 2 *This concept contains as special case precise real numbers x_0, where the characterizing function $\xi(\cdot)$ is the one-point indicator function $I_{\{x_0\}}(\cdot)$. For data in form of intervals, i. e. $x^\star = [\underline{x}, \overline{x}]$ the characterizing function $\xi(\cdot)$ is the indicator function $I_{[\underline{x},\overline{x}]}(\cdot)$ of the interval $[\underline{x}, \overline{x}]$.*

Figure 1: Characterizing function



Lemma 3 *The characterizing function $\xi(\cdot)$ of a fuzzy number $x^\star$ and its δ-cuts $C_\delta\left[x^\star\right]$ are related in the following way:*

$$\xi(x) = \max\left\{\delta \cdot I_{C_\delta[x^\star]}(x) : \delta \in [0,1]\right\} \;\forall\, x \in \mathbb{R} \tag{1}$$

Proof: *This is a special case of the representation of membership functions of fuzzy sets by their δ-cuts. Compare [5].*

3 Fuzzy vectors

Analogous to the one-dimensional case also multivariate data are frequently not precise vectors.

Examples are positions of objects on radar screens, results of imaging procedures in diagnostics, and data from remote sensing.

In case of multivariate data a generalization of real vectors is necessary to describe the imprecision of such data. This is possible by using so-called *fuzzy vectors* $\underline{x}^\star$.

Definition 4 *A fuzzy k-dimensional vector $\underline{x}^\star$ is defined by its so-called* vector-characterizing function *$\zeta(\cdot,\cdots,\cdot)$, which is a real valued function of k real variables $x_1,\cdots,x_k$ obeying the following:*

- $0 \leq \zeta(x_1,\cdots,x_k) \leq 1$
  $\forall\,(x_1,\cdots,x_k) \in \mathbb{R}^k$
- the support of $\zeta(\cdot,\cdots.\cdot)$ is a bounded subset of $\mathbb{R}^k$
- for all $\delta \in (0,1]$ the so-called δ-cut
  $C_\delta[\underline{x}^\star] := \Big\{(x_1,\cdots,x_k) \in \mathbb{R}^k :$
  $\zeta(x_1,\cdots,x_k) \geq \delta\}$
  is a non-empty finite union of simply connected compact subsets $C_{\delta,j} \subseteq \mathbb{R}^k$, for $j = 1(1)m_\delta$, i. e.

  $$C_\delta\left[\underline{x}^\star\right] = \bigcup_{j=1}^{m_\delta} C_{\delta,j},$$

  where $C_\delta\left[\underline{x}^\star\right]$ can be a singleton also.

Remark 5 *The concept of vector-characterizing functions contains also precise k-dimensional vectors $\underline{x}_0 \in \mathbb{R}^k$ by taking the indicator function $I_{\{\underline{x}_0\}}(\cdot,\cdots,\cdot)$ as vector-characterizing function.*

For fuzzy vectors $\underline{x}^\star$ and its vector-characterizing functions $\zeta(\cdot,\cdots,\cdot)$ the following is valid.

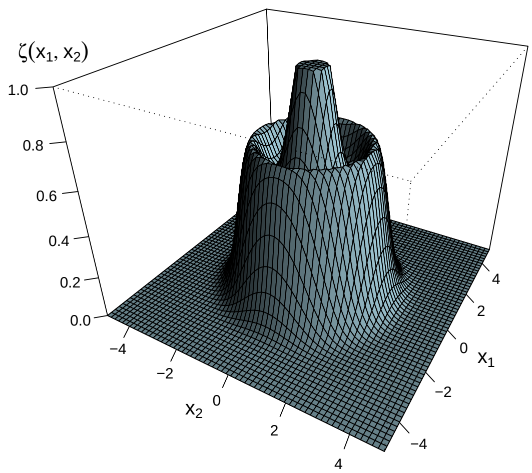
Lemma 6 *For the vector-characterizing function $\zeta(\cdot,\cdots,\cdot)$ of a non-precise vector $\underline{x}^\star$ and its δ-cuts $C_\delta\left[\underline{x}^\star\right]$ the following holds true:*

$$\begin{aligned}\zeta(x_1,\cdots,x_k) \;&=\; \max\Big\{\delta \cdot I_{C_\delta[\underline{x}^\star]}(x_1,\cdots,x_k) : \\ &\delta \in [0,1]\} \quad \forall \;\; (x_1,\cdots,x_k) \in \mathbb{R}^k\end{aligned} \tag{2}$$

The proof is analogous to that of lemma 3.

In figure 2 an example of a vector-characterizing function of a two-dimensional fuzzy vector is depicted.

Figure 2: Vector-characterizing function



4 Representation of fuzzy data in databases

For applied databases it is necessary to be able to store also fuzzy numbers and fuzzy vectors in order to provide realistic information concerning real data.

Fuzzy numbers and fuzzy vectors can be represented in databases by storing δ-cuts. This is justified by lemma 3 and lemma 6.

Remark 7 *By lemma 3 a fuzzy number $x^\star$ can be represented by the family*

$$\Big(C_\delta[x^\star];\delta \in (0,1]\Big)$$

of its δ-cuts $C_\delta\left[x^\star\right]$.

In applications a finite number of δ-cuts can be stored in databases. Depending on the necessary accuracy the number of δ-levels has to be chosen.

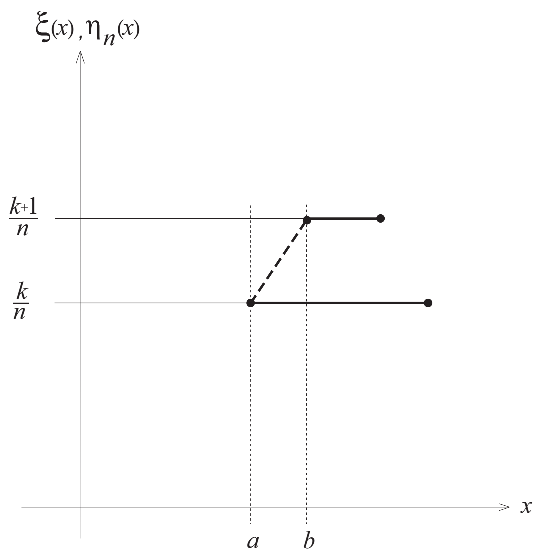
Lemma 8 *If n equidistant δ-cuts with $\delta = k/n$, $k = 1(1)n$ are stored, then the following is valid for the approximation of $\xi(\cdot)$ by the polygonal interpolation $\eta_n(\cdot)$ of the end-points of the δ-cuts:*

$$\sup\Big\{\,|\xi(x) - \eta_n(x)| : x \in \mathbb{R}\Big\} \leq \frac{1}{n}$$

ISBN: 978-1-61804-023-7

Proof: Looking at the δ-cuts for $\delta = \frac{k}{n}$ and $\delta = \frac{k+1}{n}$ the least upper bound for the distance between $\xi(x)$ and $\eta_n(x)$ is depicted in figure 3.

Figure 3: Polygonal approximation of characterizing functions



All possibilities for values $\xi(x)$ with $x \in [a, b]$ are in $\left[\frac{k}{n}; \frac{k+1}{n}\right]$. Therefore we obtain
$|\xi(x) - \eta_n(x)| \leq \frac{1}{n}$.

Remark 9 *Also fuzzy multivariate data can be represented in databases by storing a suitable family of δ-cuts of the corresponding vector-characterizing function.*

5 Conclusion

The design of future databases has to take care of the fact, that many real data are fuzzy. In order to provide uncertainty based information it is necessary to construct so-called *uncertainty based databases* which are able to store fuzzy information in form of fuzzy numbers and fuzzy vectors respectively.

References:

[1] J. Galindo: *Handbook of Research on Fuzzy Information Processing in Data Bases*, IGI Publishing, Hershey, 2008

[2] G. Klir, B. Yuan: *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, NJ, 1995

[3] F. Petry: *Fuzzy Databases - Principles and Applications*, Kluwer, Boston, 1996

[4] R. Viertl: On the Description and Analysis of Measurements of Continuous Quantities, *Kybernetika*, Vol. 38 (2002)

[5] R. Viertl: *Statistical Methods for Fuzzy Data*, Wiley, Chichester, 2011

[6] O. Wolkenhauer: *Data Engineering*, Wiley, New York, 2001

ISBN: 978-1-61804-023-7