STATISTICAL INFERENCE WITH IMPRECISE DATA

Reinhard Viertl
Vienna University of Technology, Wien, Austria

Keywords: Bayesian statistics, Classical statistics, Data quality, Fuzzy information, Fuzzy numbers, Fuzzy probability, Imprecision, Imprecise data, Standard statistics, Statistical estimation, Statistics

Contents

Summary

Real measurement results of continuous quantities are not precise real numbers but more or less non-precise. This kind of uncertainty is different from measurement errors. Contrary to standard statistical inference, where the data are assumed to be numbers or vectors, for imprecise data statistical methods have to be generalized in order to be able to analyze such data. Generalized inference procedures for imprecise data are given in the article.

1. Imprecise data

The results of measurements are not precise numbers or vectors but more or less imprecise numbers or vectors. This uncertainty is different from measurement errors and stochastic uncertainty and is called *imprecision.* Imprecision is a feature of single observations form continuous quantities. Errors are described by statistical models and should not be confused with imprecision. In general imprecision and errors are superimposed. *In this article errors are not considered.*

Example: Many measurements in environmetrics are connected with a remarkable amount of uncertainty and especially imprecision. For example the data on the concentration of toxic substances in different environmental media are imprecise quantities and their measurements are not precise. The same is true for total amounts of dangerous substances released to the environment.

Example: The life time of a system can in general not be described by one real number because the time of the end of the life time- for example a tree - is not a precise number but more or less non-precise.

Example: The results of many observations are color intensity pictures (for example observations from remote sensing). The resulting data are not precise numbers but imprecise numbers or imprecise vectors.

A special case of imprecise data are *interval data*.

Precise real numbers $x_0 \in \mathbb{R}$ as well as intervals $[a,b] \subseteq \mathbb{R}$ are uniquely characterized by their *indicator functions* $I_{\{x_0\}}(\cdot)$ and $I_{[a,b]}(\cdot)$ respectively, where the indicator function $I_A(\cdot)$ of a classical set A is defined by

$$I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A. \end{cases} \tag{1}$$

Often the imprecision of measurements implies that exact boundaries of interval data are not realistic. Therefore it is necessary to generalize real numbers and intervals to describe imprecision. This is done by the concept of imprecise *numbers* as generalization of real numbers and intervals. Imprecise numbers as well as imprecise subsets of $\mathbb{R}$ are described by generalizations of indicator functions, called *characterizing functions*.

2. Imprecise numbers and Characterizing Functions

To describe mathematically imprecise observations imprecise *numbers* $x^\star$ are modelled by so-called *characterizing functions* $\xi(\cdot)$, which characterize the imprecision of a single observation.

Definition 1: A *characterizing function* $\xi(\cdot)$ of imprecise number or an imprecise interval is a real function of a real variable with the following properties:

a) $\xi : \mathbb{R} \to [0,1]$

b) $\exists\, x_0 \in \mathbb{R} : \ \xi(x_0) = 1$

c) $\forall\, \alpha \in (0,1]$ the set $B_\alpha := \{x \in \mathbb{R} : \xi(x) \geq \alpha\} = [a_\alpha, b_\alpha]$ is a finite closed interval, called α -*cut* of $\xi(\cdot)$.

The set *supp*$(\xi(.)) := \{x \in \mathbb{R} : \xi(x) > 0\}$ is called *support* of $\xi(\cdot)$.

In Figure 1 some examples of characterizing functions are depicted.

In Figure 2 one α-cut of a characterizing function is explained.

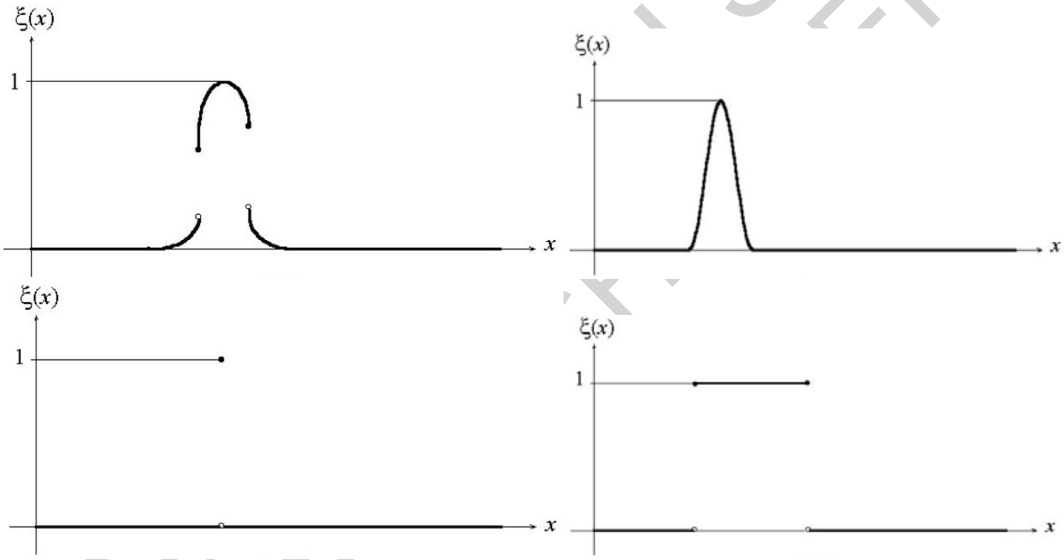

Figure 1: Examples of characterizing functions

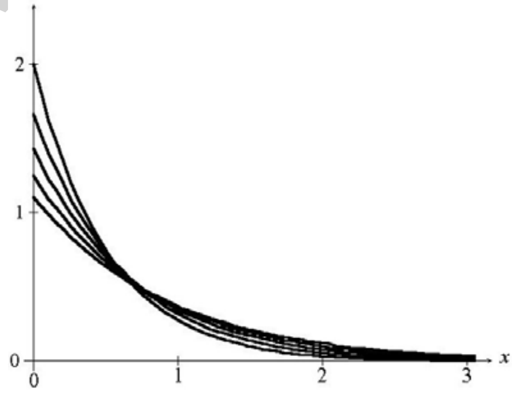

Figure 2: α-cut of a characterizing function

Remark: Imprecise observations and imprecise numbers will be marked by stars, i.e. $x^\star$, to distinguish them from (precise) real numbers x. Every imprecise number $x^\star$ is characterized by the corresponding characterizing function $\xi_{x^\star}(\cdot)$. The set of all imprecise numbers is denoted by $\mathcal{N}(\mathbb{R})$.

Remark: In fuzzy set theory a less specific analogue to the characterizing function is called *membership function*.

Remark: Imprecise numbers and imprecise intervals can be considered as special fuzzy subsets of $\mathbb{R}$. Therefore they are reasonable generalizations of real numbers and intervals, because precise numbers as well as intervals are classical subsets of $\mathbb{R}$.

Proposition 1: Characterizing functions $\xi(\cdot)$ are uniquely determined by the family

$$\left(B_\alpha ; \alpha \in (0,1]\right)$$

of their α-cuts B_α and the following holds

$$\xi(x) = \max_{\alpha \in (0,1]} \alpha \cdot I_{B_\alpha}(x) \quad \forall x \in \mathbb{R}. \tag{2}$$

Proof: Let $x_0 \in \mathbb{R}$ then it follows

$$\begin{aligned} \alpha \cdot I_{B_\alpha}(x_0) &= \alpha \cdot I_{\{x:\xi(x)\geq\alpha\}}(x_0) \\ &= \begin{cases} \alpha & \text{for} \quad \xi(x_0) \geq \alpha \\ 0 & \text{for} \quad \xi(x_0) < \alpha \end{cases} \end{aligned}$$

and from that $\alpha \cdot I_{B_\alpha}(x_0) \leq \xi(x_0) \quad \forall \alpha \in (0,1]$
and $\sup\limits_{\alpha \in (0,1]} \alpha \cdot I_{B_\alpha}(x_0) \leq \xi(x_0)$.

For $\alpha_0 = \xi(x_0)$ we obtain $B_{\alpha_0} = \{x : \xi(x) \geq \xi(x_0)\} = \left[a_{\alpha_0}, b_{\alpha_0}\right]$ and therefore

$$\begin{aligned} \alpha_0 \cdot I_{B_\alpha}(x_0) &= \alpha_0 \cdot 1 = \xi(x_0) \\ &= \max_{\alpha \in (0,1]} \alpha \cdot I_{B_\alpha}(x_0). \end{aligned}$$

2.1. Special Imprecise Numbers

The characterizing function $\xi_{x^\star}(\cdot)$ of an imprecise number $x^\star$ or an imprecise interval can be written in the following way

$$\xi_{x^\star}(x) = \begin{cases} L(x) & \text{for} \quad x \le m_1 \\ 1 & \text{for} \quad m_1 \le x \le m_2, \\ & \qquad\qquad \text{with } m_1 \le m_2 \\ R(x) & \text{for} \quad x \ge m_2 \end{cases} \tag{3}$$

where $L(\cdot)$ is an increasing real function and $R(\cdot)$ a decreasing real function with

$$\lim_{x \downarrow -\infty} L(x) = 0 \quad \text{and} \quad \lim_{x \uparrow \infty} R(x) = 0. \tag{4}$$

The following special forms of characterizing functions are frequently used:

Trapezoidal imprecise numbers $t^\star(m_1, m_2, a_1, a_2)$:

$$L(x) = \max\left(0, \ \frac{x - m_1 + a_1}{a_1}\right) \tag{5}$$

$$R(x) = \max\left(0, \frac{m_2 + a_2 - x}{a_2}\right) \tag{6}$$

In Figure 3 characterizing functions of trapezoidal imprecise numbers are depicted.

For $m_1 = m_2 = m$ so called *triangular* imprecise *numbers* $t^\star(m, a_1, a_2)$ are obtained.

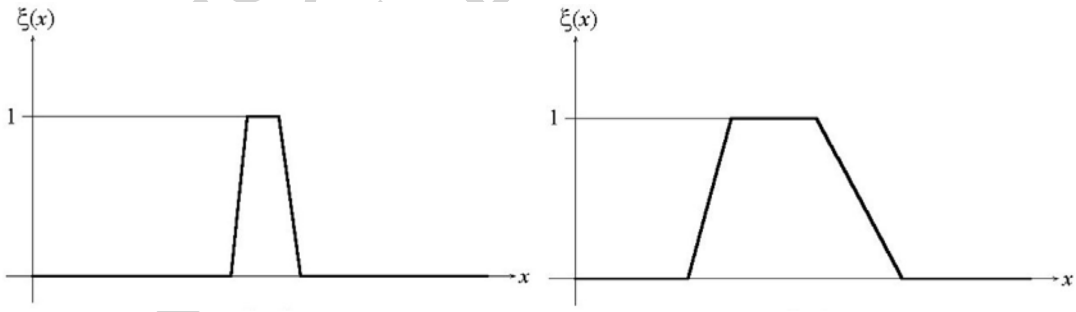

Figure 3: Trapezoidal imprecise numbers, also called imprecise intervals

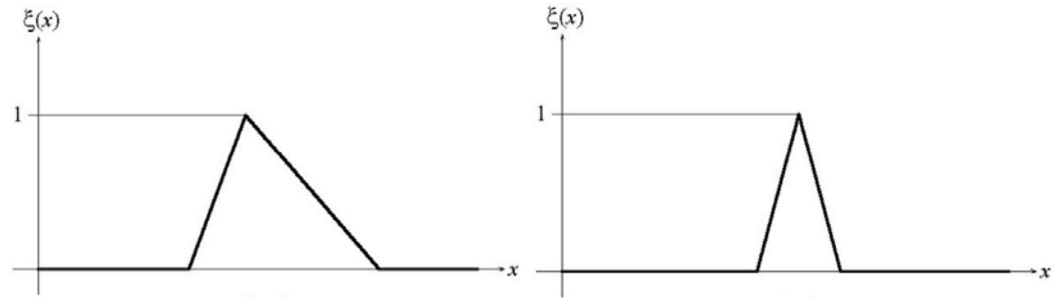

Figure 4: Triangular imprecise numbers.

In Figure 4 there are two examples of triangular imprecise numbers.

Example: The measurement of the concentration of a poison in the air at a fixed location and fixed time can be described by a trapezoidal imprecise number.

The problem of how to obtain the characterizing function of the imprecise number describing the imprecise quantity *water level* is discussed in Section 3.

2.2. Convex Hull of a Non-convex Pseudo-characterizing Function

For continuous functions $\varphi(\cdot)$ which fulfill conditions (a) and (b) of definition 1 but not condition (c), the function $\varphi(\cdot)$ can be transformed into a characterizing function $\xi(\cdot)$ fulfilling also condition (c):

Definition 2: Let $\varphi(\cdot)$ be a real continuous function fulfilling conditions (a) and (b) of definition 1 but not condition (c), such that some α-cuts B_α of $\varphi(\cdot)$ are unions of disjoint intervals $B_{\alpha,i,}$ i.e. $B_\alpha = \bigcup_{i=1}^{k_\alpha} B_{\alpha,i}$ with $B_{\alpha,i} = \left[a_{\alpha,i}\,;b_{\alpha,i}\right]$. Using proposition 1 the so called *convex hull* $\xi(\cdot)$ of $\varphi(\cdot)$ is the characterizing function defined via its α-cuts $C_\alpha\left(\xi(\cdot)\right)$ by

$$C_\alpha := \left[\min_{i=1(1)k_\alpha} a_{\alpha,i}\,;\,\max_{i=1(1)k_\alpha} b_{\alpha,i}\right]. \tag{7}$$

By proposition 1, $\xi(\cdot)$ is given by its value $\xi(x) = \max_{\alpha\in(0,1]} \alpha \cdot I_{C_\alpha}(x) \quad \forall\, x \in \mathbb{R}$.

The concept of convex hull is explained in Figure 5.

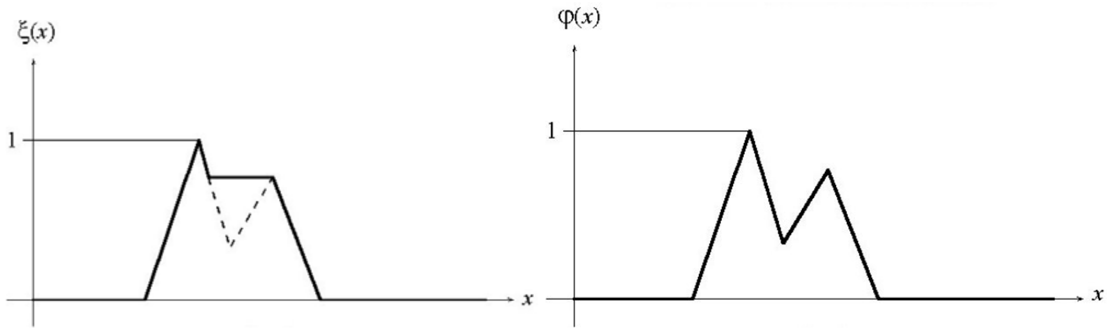

Figure 5: Pseudo-characterizing function $\varphi(\cdot)$ and its convex hull $\xi(\cdot)$

-
-
-

Bibliography

Bandemer H. (Ed.): *Modelling Uncertain Data*, Akademie Verlag, Berlin, 1993. [Contains different aspects of uncertainty analysis]

Dubois D., Prade H.: Fuzzy sets and statistical data, *European Journal of Operational Research* 25, 345-356 (1986) [Basic ideas of fuzzy concepts in statistical context]

Frühwirth-Schnatter S.: On fuzzy Bayesian inference, *Fuzzy Sets and Systems* 60 (1993). [One of the first papers on fuzzy methods in Bayesian inference]

Gil M.A., Corral N., Gil P.: The minimum inaccuracy estimates in χ^2 tests for goodness of fit with fuzzy observations, *Journal of Statistical Planning and Inference* 19, 95-115 (1988). [Early paper on statistical tests and imprecise data]

Kacprzyk J., Fedrizzi M. (Eds.): *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making*, Lecture Notes in Economics and Mathematical Systems, Vol.310, Springer-Verlag, Berlin, 1988. [Contains papers on imprecise information in the context of decisions]

Keresztfalvi T.: Operations on Fuzzy Numbers Extended by Yager's Family of t-Norms, in: H. Bandemer (Ed.): *Modelling Uncertain Data*, Akademie Verlag, Berlin, 1993. [Reference for general combination rules]

Kruse R., Meyer K.D.: *Statistics with Vague Data*, D. Reidel Publ., Dordrecht, 1987. [More mathematical models for imprecise data and statistics]

Niculescu S.P., Viertl R.: A comparison between two fuzzy estimators for the mean, *Fuzzy Sets and Systems* 48,341-350 (1992). [Generalized estimation method based on imprecise data]

Viertl R.: Is it necessary to develop a fuzzy Bayesian inference, in: R. Viertl (Ed.): *Probability and Bayesian Statistics*, Plenum Press, New York, 1987. [Early paper on imprecise information and Bayesian statistics]

Viertl R.: Statistical inference for fuzzy data in environmetrics, *Environmetrics* 1, 37-42 (1990). [Relevance of imprecision for environmental data]

Viertl R.: *Statistics with Non-Precise Data*, CRC Press, Boca Raton, Florida, 1996. [Basic text on the description and statistical analysis of imprecise data]

Viertl R.: On Fuzzy Predictive Densities, *Tatra Mountains Mathem*. *Publications* 16 (1999) [Recent result on predictions based on imprecise information]

Biographical Sketch

Reinhard Viertl Born March 25, 1946, at Hall in Tyrol, Austria. Studies in civil engineering and engineering mathematics at the Technische Hochschule Wien. Receiving a Dipl.-Ing. degree in engineering mathematics in 1972. Dissertation in mathematics and Doctor of engineering science degree in 1974. Appointed assistant at the Technische Hochschule Wien and promotion to University Docent in 1979. Research fellow and visiting lecturer at the University of California, Berkeley, from 1980 to 1981, and visiting Docent at the University of Klagenfurt, Austria in winter 1981 - 1982. Since 1982 full professor of applied statistics at the Department of Statistics, Vienna University of Technology. Visiting professor at the Department of Statistics, University of Innsbruck, Austria from 1991 to 1993. He is a fellow of the Royal Statistical Society, London, held the Max Kade fellowship in 1980, and is founder of the Austrian Bayes Society, member of the International Statistical Institute, president of the Austrian Statistical Society from 1987 to 1995. Invitation to membership in the New York Academy of Sciences in 1998. Author of the books *Statistical Methods in Accelerated Life Testing* (1988), *Introduction to Stochastics* in German language (1990), *Statistical Methods for* Imprecise *Data* (1996). Editor of the books *Probability and Bayesian Statistics* (1987), *Contributions to Environmental Statistics* in German language (1992). Co-editor of a book titled *Mathematical and Statistical Methods in Artificial Intelligence* (1995), and co-editor of two special volumes of journals. Author of over 70 scientific papers in algebra, probability theory, accelerated life testing, regional statistics, and statistics with imprecise data. Editor of the publication series of the Vienna University of Technology, member of the editorial board of scientific journals, organiser of different scientific conferences.