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

1. Imprecise data
2. Imprecise numbers and characterizing functions  
2.1. Special Imprecise Numbers  
2.2. Convex Hull of a Non-convex Pseudo-characterizing Function  
3. Construction of characterizing functions  
4. Multivariate data, imprecise vectors, and combination of imprecise samples  
4.1. Imprecise multivariate data and imprecise vectors  
4.2. Combination of Imprecise Observations  
4.3. Multivariate Data  
5. Functions and imprecision  
5.1. Functions of Imprecise Variables  
5.2. Functions with Imprecise Values  
6. Generalized inference procedures for imprecise samples  
7. Classical statistical inference for imprecise data  
7.1. Generalized Point Estimators for Parameters  
7.2. Generalized Confidence Regions for Parameters  
7.3. Statistical Tests based on Imprecise Data  
8. Bayesian inference for imprecise data  
8.1. Bayes’ Theorem for Imprecise Data  
8.2. Bayes Estimators  
8.3. Generalized HPD-regions  
8.4. Fuzzy Predictive Distributions  
8.5. Fuzzy a priori Distributions  
9. Conclusion  
Glossary  
Bibliography  
Biographical Sketch

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

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

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 $x^*$ 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 \( \alpha \)-cut of \( \xi(\cdot) \).

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

In Figure 1 some examples of characterizing functions are depicted.

In Figure 2 one \( \alpha \)-cut of a characterizing function is explained.

Figure 1: Examples of characterizing functions

Figure 2: \( \alpha \)-cut of a characterizing function
**Remark:** Imprecise observations and imprecise numbers will be marked by stars, i.e. $x^*$, to distinguish them from (precise) real numbers $x$. Every imprecise number $x^*$ is characterized by the corresponding characterizing function $\xi_{x^*}(\cdot)$. 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

\[
(B_\alpha; \alpha \in (0,1])
\]

of their $\alpha$-cuts $B_\alpha$ and the following holds

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

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

\[
\alpha \cdot I_{B_\alpha}(x_0) = \alpha \cdot I_{\{x: \xi(x) \geq \alpha\}}(x_0)
\]

\[
= \begin{cases} 
\alpha \text{ for } \xi(x_0) \geq \alpha \\
0 \text{ for } \xi(x_0) < \alpha 
\end{cases}
\]

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

\[
\sup_{\alpha \in (0,1]} \alpha \cdot I_{B_\alpha}(x_0) \leq \xi(x_0).
\]

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

\[
\alpha_0 \cdot I_{B_{\alpha_0}}(x_0) = \alpha_0 \cdot 1 = \xi(x_0)
\]

\[
= \max_{\alpha \in (0,1]} \alpha \cdot I_{B_\alpha}(x_0).
\]

### 2.1. Special Imprecise Numbers

The characterizing function $\xi_{x^*}(\cdot)$ of an imprecise number $x^*$ or an imprecise interval can be written in the following way
\[ \xi_x(x) = \begin{cases} 
L(x) & \text{for } x \leq m_1 \\
1 & \text{for } m_1 \leq x \leq m_2, \\
R(x) & \text{for } x \geq m_2 
\end{cases} 
\]  

(3)

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

\[
\lim_{x \to -\infty} L(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} R(x) = 0.
\]  

(4)

The following special forms of characterizing functions are frequently used:

**Trapezoidal imprecise numbers** \( t^* \left( m_1, m_2, a_1, a_2 \right) \):

\[
L(x) = \max \left( 0, \frac{x - m_1 + a_1}{a_1} \right)
\]  

(5)

\[
R(x) = \max \left( 0, \frac{m_2 + a_2 - x}{a_2} \right)
\]  

(6)

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

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

---

Figure 3: Trapezoidal imprecise numbers, also called imprecise intervals
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 \( \alpha \)-cuts \( B_\alpha \) of \( \varphi(\cdot) \) are unions of disjoint intervals \( B_{\alpha,i} \), i.e. \( B_\alpha = \bigcup_{i=1}^{k} B_{\alpha,i} \) with \( B_{\alpha,i} = [a_{\alpha,i}; b_{\alpha,i}] \). Using proposition 1 the so called **convex hull** \( \xi(\cdot) \) of \( \varphi(\cdot) \) is the characterizing function defined via its \( \alpha \)-cuts \( C_\alpha(\xi(\cdot)) \) by

\[
C_\alpha \bigg( \xi(\cdot) \bigg) \ni \min_{i=1}^{k} a_{\alpha,i} ; \max_{i=1}^{k} b_{\alpha,i}.
\]

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

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]


**Biographical Sketch**