Fuzzy Sets and Statistics

REINHARD VIERTL
Department of Statistics and Probability Theory
Vienna University of Technology
Wiedner Hauptstrasze 8-10
A-1040 Wien
AUSTRIA

Abstract: - Real Data are often not precise but more or less fuzzy. This is always the case for measurements of continuous quantities. Before making statistical analyses of such data this kind of uncertainty has to be described quantitatively. This can be done using fuzzy numbers which are characterized by so called characterizing functions. These characterizing functions are special types of membership functions. In the fuzzy set literature they are also called fuzzy intervals.

Based on this description of fuzzy data a non-precise sample consists of a finite sequence $x_1^*, \cdots, x_n^*$ of fuzzy numbers x_i^*.

Facing this kind of data statistical methods have to be adapted to the situation of fuzzy data. This is possible and related statistical procedures will be presented in the contribution. These are especially:

Descriptive statistical methods
Generalized point estimation procedures
Generalized confidence intervals
Generalized classical test procedures
Generalized Bayesian methods

Key-Words:- Bayesian inference, descriptive statistics, fuzzy data, non-precise numbers, parameter estimation, statistical inference, statistical tests

1 Non-Precise Data and Fuzzy Sets

Real data are often not precise numbers or vectors but more or less non-precise. This is true for qualitative data as well as for measurement results. Even precision measurements of continuous quantities are not precise numbers but more or less fuzzy.

In order to apply statistical procedures to non-precise data it is necessary to model non-precise data in a reasonable quantitative (mathematical) way.

In case of one-dimensional data the individual observations can be modelled by special membership functions of special fuzzy subsets of the set $I\!R$ of real numbers. These special membership functions are called *characterizing functions* because they are characterizing non-precise numbers as suitable generalizations of indicator functions $I_{\{x\}}(\cdot)$ of precise real numbers x, and indicator functions.

A characterizing function is a real function $x(\cdot)$ of a real variable x which obeys the following:

(1) $0 \le x(x) \le 1$ for all $x \in I\!R$

(2) $\exists x_0 \in I\!R : x(x_0) = 1$

(3) $\forall d \in (0,1]$ the so-called d - cut
$C_d[x(\cdot)] := \{x \in I\!R : x(x) \ge d\}$
is a closed finite interval $[a_d, b_d]$

A characterizing function characterizes a non-precise number x^* which describes a non-precise observation.

Remark: In fuzzy set theory the above defined non-precise numbers are also called fuzzy intervals.

Examples of one-dimensional non-precise data are geodetical distances, life times of biological units, and weights.

An important topic is how to obtain the characterizing function of a non-precise datum x^*. Details on this are given in the monograph [7].

For vector valued quantities a generalization of real vectors is necessary. This is possible defining so-called *non-precise vectors*. A non-precise vector is defined by its so-called *vector-characterizing function* $z(\cdot, \cdots, \cdot)$ which is a real function of k real variables obeying the following:

(I) $0 \le z(x_1, \cdots, x_k) \le 1$ for all $(x_1, \cdots, x_k) \in I\!R^k$

(II) $\exists (x_1, \cdots, x_k) \in I\!R^k : z(x_1, \cdots, x_k) = 1$

(III) $\forall d \in (0,1]$ the so-called d - cut

$$C_d[z(\cdot,\cdots,\cdot)] := \{(x_1,\cdots,x_k) \in \mathbb{R}^k : z(x_1,\cdots,x_k) \geq d\}$$

is a compact and convex subset of $\mathbb{R}^k$

Remark: Vector-characterizing functions are special membership functions of fuzzy subsets of the *k*-dimensional Euclidian space $\mathbb{R}^k$.

For a precise *k*-dimensional vector $\underline{x} = (x_1,\cdots,x_k)$ the corresponding vector-characterizing function is the one-point indicator function $I_{\{(x_1,\cdots,x_k)\}}(\cdot,\cdots,\cdot)$. For a *k*-dimensional interval $\underset{i=1}{\overset{k}{\mathrm{X}}}[a_i,b_i]$ the corresponding vector-characterizing function is the indicator function $I_{\underset{i=1}{\overset{k}{\mathrm{X}}}[a_i,b_i]}(\cdot,\cdots,\cdot)$.

An example of how to obtain the vector-characterizing function of a non-precise vector is given for a 2-dimensional quantity by the following: Let the position of an object be given on a radar screen. The light intensity determines the vector-characterizing function $z(\cdot,\cdot)$ of the non-precise position vector $\underline{x}^*$. Let $f(x_1,x_2)$ denote the light intensity in the plane $\mathbb{R}^2$. Then the vector-characterizing function $z(\cdot,\cdot)$ of the 2-dimensional non-precise vector $\underline{x}^*$ is given by its values

$$z(x_1,x_2) = \frac{f(x_1,x_2)}{\max\limits_{(x_1,x_2)\in\mathbb{R}^2} f(x_1,x_2)} \quad \text{for all } (x_1,x_2) \in \mathbb{R}^2.$$

2 Non-Precise Samples

A sample of a stochastic quantity *X* consists of *n* observations $x_1,\cdots,x_n$. In case of non-precise observations – which is always the case for continuous quantities – the sample consists of *n* non-precise numbers $x_1^*,\cdots,x_n^*$ with corresponding characterizing functions $x_1(\cdot),\cdots,x_n(\cdot)$.

In statistical inference the elements x_i of the observation space M_X of the stochastic quantity *X* are combined to a vector $\underline{x} = (x_1,\cdots,x_n)$ which is an element of the *sample space* M_X^n with

$$M_X^n = M_X \times M_X \times \cdots \times M_X,$$

i. e. the Cartesian product of *n* copies of the observation space M_X. Statistical procedures are functions defined on the sample space.

In the standard setting of statistics the combination of a sample is trivial. This is not so for non-precise samples by the following reason: A non-precise sample consists of a vector $(x_1^*,\cdots,x_n^*)$ of non-precise numbers $x_1^*,\cdots,x_n^*$. This is essentially different to a non-precise vector $\underline{x}^*$ in the sample space M_X^n. But in order to generalize statistical procedures using the *extension principle*, the non-precise sample has to be combined into a non-precise vector. This is done by a so-called *combination rule* which is a family of functions related to triangular norms.

The vector-characterizing function $z(\cdot,\cdots,\cdot)$ of the *combined non-precise sample* $\underline{x}^*$ is given, using the characterizing functions $x_i(\cdot)$ of x_i^*, by a *combination rule* C which gives a combination $C_n(\cdot,\cdots,\cdot)$ for every $n \in \mathbb{N}$:

$$z(x_1,\cdots,x_n) = C_n(x_1(x_1),\cdots,x_n(x_n))$$

for all $(x_1,\cdots,x_n) \in \mathbb{R}^n$.

The most important combination rule is

$$z(x_1,\cdots,x_n) = \min_{i=1(1)n} x_i(x_i) \quad \text{for all } (x_1,\cdots,x_n) \in \mathbb{R}^n.$$

In general combination rules have to fulfill the following:

(i) $C_1[x_1(x)] = x_1(x)$ for all $x \in \mathbb{R}$

(ii) $C_n[I_{\{\bar{x}_1\}}(x_1),\cdots,I_{\{\bar{x}_n\}}(x_n)] = I_{\{(\bar{x}_1,\cdots,\bar{x}_n)\}}(x_1,\cdots,x_n)$

for all precise $\bar{x}_i \in \mathbb{R}$ and all $(x_1,\cdots,x_n) \in \mathbb{R}^n$

(iii) $C_n[I_{[a_1,b_1]}(x_1),\cdots,I_{[a_n,b_n]}(x_n)] = I_{[a_1,b_1]\times\cdots\times[a_n,b_n]}(x_1,\cdots,x_n)$

for all intervals $[a_i,b_i]$ and all $(x_1,\cdots,x_n) \in \mathbb{R}^n$

3 Descriptive Statistics for Non-Precise Data

Empirical distribution functions as well as histograms have to be adapted to the situation of non-precise data. This can be done in the following way: The characterizing functions $x_i(\cdot)$ of the non-precise observations x_i^* are used to construct a histogram whose heights over the classes are non-precise numbers. For details see [6]. For generalizations of the empirical distribution function see the monograph [7].

4 Generalizations of Statistical Inference

Statistical inference procedures can be adapted to the situation of non-precise data using the combined non-precise sample $\underline{x}^*$ and the extension principle from fuzzy set theory.

Let $z(\cdot,\cdots,\cdot)$ be the vector-characterizing function of the combined non-precise sample $\underline{x}^*$. Then the generalized (fuzzy) value

$$s^* = \mathrm{S}(x_1^*,\cdots,x_n^*) = \mathrm{S}(\underline{x}^*)$$

of a classical real valued statistic

$$S = \mathrm{S}(X_1,\cdots,X_n),$$

i. e. $S: M_X^n \to I\!R$ with continuous function $S(\cdot,\cdots,\cdot)$ is given by the characterizing function $h(\cdot)$ of s^*, whose values $h(y)$ are determined by the extension principle to be

$$h(y) = \begin{Bmatrix} \sup\{z(\underline{x}): S(\underline{x}) = y\} \text{ if } \exists \underline{x}: S(\underline{x}) = y \\ 0 \qquad\qquad \text{otherwise} \end{Bmatrix} \forall\, y \in I\!R .$$

Remark: The continuity of $S(\cdot,\cdots,\cdot)$ guarantees that $h(\cdot)$ is a characterizing function in the sense of the definition given in section 1.

4.1 Parameter Estimation

Point estimators $\hat{q} = J(x_1,\cdots,x_n)$ for parameters q can be adapted by the definition above immediately. Here $J: M_X^n \to \Theta$ is a function from the sample space to the parameter space Θ of the stochastic model $X \sim f(\cdot|q), q \in \Theta$. The fuzzy value of the generalized estimator

$$\hat{q}^* = J(\underline{x}^*)$$

is given by its characterizing function $h(\cdot)$ as above.

The generalization of confidence sets is possible in the following way: Let $k(X_1,\cdots,X_n)$ be a confidence function with given confidence level $1-\alpha$, i. e. for precise data $x_1,\cdots,x_n$ the value $k(x_1,\cdots,x_n)$ is a subset of Θ with probability of covering the true parameter q_0 is $1-\alpha$, i. e.

$$Pr\{q_0 \in k(X_1,\cdots,X_n)\} = 1-\alpha ,$$

then the generalized confidence set based on non-precise data with non-precise combined sample is a fuzzy subset Θ^* of the parameter space Θ. The membership function $j(\cdot)$ of Θ^* is defined by

$$j(\cdot) = \begin{Bmatrix} \sup\{z(\underline{x}): q \in k(\underline{x})\} \text{ if } \exists \underline{x}: q \in k(\underline{x}) \\ 0 \qquad\qquad \text{otherwise} \end{Bmatrix}$$

for all $q \in \Theta$.

Remark: Generalized confidence sets are classical examples of fuzzy subsets. For precise data the given concept yield the indicator function of the classical confidence set.

4.2 Statistical Tests

Classical test statistics $T = t(X_1,\cdots,X_n)$ yield precise values $t = t(x_1,\cdots,x_n)$ for precise sample $x_1,\cdots,x_n$. In case of fuzzy samples $x_1^*,\cdots,x_n^*$ the values of test statistics become fuzzy. Therefore suitable decision rules for acceptance or rejection of hypotheses are necessary. This is possible using the concept of *p*-values. For details see the forthcoming encyclopedia article [10].

4.3 Bayesian Inference

Adapting Bayes' theorem

$$p(q \mid x_1,\cdots,x_n) \propto p(q) \cdot \ell(q; x_1,\cdots,x_n) \text{ for } q \in \Theta$$

to the situation of non-precise data $x_1^*,\cdots,x_n^*$ an a-posteriori density $p^*(q \mid x_1^*,\cdots,x_n^*)$ with fuzzy values of the density is obtained. Based on this fuzzy a-posteriori density predictions and decisions are possible. Details on this are given in the book [7] and in the paper [8] .

References:

[1] H. Bandemer: *Ratschläge zum mathematischen Umgang mit Ungewißheit - Reasonable Computing*, Teubner Verlag, Stuttgart, 1997

[2] H. Bandemer, W. Näther: *Fuzzy Data Analysis*, Kluwer, Dordrecht, 1992

[3] S. Bodjanova: A generalized histogram, *Fuzzy Sets and Systems* (2000)

[4] K. Menger: Ensembles flous et fonctions aléatoires, *Comptes Rendus Acad. Sci.* 232 (1951)

[5] R. Seising (Ed.): *Fuzzy Theorie und Stochastik*, Vieweg, Braunschweig, 1999

[6] R. Viertl: Statistics with Non-Precise Data, *Journal of Computing and Information Technology*, Vol. 4 (1996))

[7] R. Viertl: *Statistical Methods for Non-Precise Data*, CRC Press, Boca Raton, Florida, 1996

[8] R. Viertl: On Fuzzy Predictive Densities, *Tatra Mountains Mathem. Publications* 16 (1999)

[9] R. Viertl: Fuzzy Numbers and Non-Precise Data, in: *Encyclopedia of Environmetrics*, Wiley, New York, 2001

[10] R. Viertl: Statistical Inference for Non-Precise Data, in: *Encyclopedia of Life Support Systems*, UNESCO, Paris, (to appear)