Metrika (2004) 59: 21–29
DOI 10.1007/s001840300269

Metrika

Testing hypotheses with fuzzy data: The fuzzy *p*-value

P. Filzmoser and R. Viertl

Department of Statistics and Probability Theory, Vienna University of Technology, Wiedner Hauptstraße 8–10, A-1040 Wien, Austria (E-mail: P.Filzmoser@tuwien.ac.at)

Abstract. Statistical hypothesis testing is very important for finding decisions in practical problems. Usually, the underlying data are assumed to be precise numbers, but it is much more realistic in general to consider fuzzy values which are non-precise numbers. In this case the test statistic will also yield a non-precise number. This article presents an approach for statistical testing at the basis of fuzzy values by introducing the *fuzzy p-value*. It turns out that clear decisions can be made outside a certain interval which is determined by the characterizing function of the fuzzy *p*-values.

Key words: Fuzzy data, Non-precise numbers, *p*-value, Hypothesis testing

1 Introduction

Real observations of continuous quantities are not precise numbers but more or less non-precise. The best description of such data is by so-called non-precise numbers. Such observations are also called fuzzy. The fuzziness is different from measurement errors and stochastic uncertainty. It is a feature of single observations from continuous quantities. Errors are described by statistical models and should not be confused with fuzziness. In general fuzziness and errors are superimposed.

A typical example for a non-precise number is the life time of a system which can in general not be described by one real number because the time of the end of the life time is not a precise number but more or less non-precise. Other examples of non-precise data are data given by color intensity pictures or readings on an analogue measurement equipment. Also readings on digital measurement equipments are not precise numbers but intervals since there is only a finite number of decimals available. Further examples are given in Viertl (2002).

A special case of non-precise data are data in form of intervals. 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

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

Often the fuzziness of measurements implies that exact boundaries of interval data are not realistic. Therefore it is necessary to generalize real numbers and intervals to describe fuzziness. This is done by the concept of *non-precise numbers* as generalization of real numbers and intervals. Non-precise numbers as well as non-precise subsets of $\mathbb{R}$ are described by generalizations of indicator functions, called *characterizing functions* $\xi(\cdot)$ (see, e.g., Viertl, 1996). Characterizing functions are real functions $\xi : \mathbb{R} \longrightarrow [0, 1]$ with the following properties:

(i) $0 \leq \xi(x) \leq 1 \quad \text{for all} \quad x \in \mathbb{R}$
(ii) $\exists\, x_0 \in \mathbb{R} : \xi(x_0) = 1$
(iii) For all $\delta \in (0, 1]$ the so-called δ-cut $C_\delta(\xi(\cdot)) := \{x \in \mathbb{R} : \xi(x) \geq \delta\} = [a_\delta, b_\delta]$ is a closed finite interval.

In the following, non-precise observations and non-precise numbers will be marked by stars, i.e. x^*, to distinguish them from (precise) real numbers x. In order to simplify the notation we denote the δ-cut of a characterizing function $\xi(\cdot)$ of a fuzzy observation x^* by $C_\delta(x^*)$.

Methods how to obtain the characterizing function of one-dimensional non-precise observations are given in Viertl (2002).

Remark. *The concept of non-precise numbers is more general than the concept of fuzzy numbers. It contains both fuzzy numbers and fuzzy intervals. This is necessary because real fuzzy data are of this more general shape.*

The paper is organized as follows. In Section 2 we explain how the vector of fuzzy observations is combined to form a non-precise element of the sample space. Section 3 is concerned with testing hypotheses with precise numbers. In Section 4 we extend the testing problem to fuzzy values of a test statistic and introduce the fuzzy p-value. This concept is illustrated in several examples with typical testing problems (Section 5). In the final section some conclusions and proposals for further research are given.

2 Statistics with fuzzy data

Let us consider a univariate random variable X. Drawing a sample of n observations of X results in n non-precise numbers $x_1^\star, \ldots, x_n^\star$. In a general setting, these n non-precise numbers will have different characterizing functions denoted by $\xi_1(\cdot), \ldots, \xi_n(\cdot)$. We can combine these non-precise numbers into an n-dimensional fuzzy vector $\boldsymbol{x}^\star$ which is determined by a so-called *vector characterizing function* $\zeta(\cdot, \ldots, \cdot)$. The function $\zeta : \mathbb{R}^n \longrightarrow [0, 1]$ has the following properties:

(i) $0 \leq \zeta(x_1, \ldots, x_n) = \zeta(\boldsymbol{x}) \leq 1 \quad \text{for all} \quad (x_1, \ldots, x_n) = \boldsymbol{x} \in \mathbb{R}^n$
(ii) $\exists\, \boldsymbol{x}_0 \in \mathbb{R}^n : \zeta(\boldsymbol{x}_0) = 1$
(iii) $\forall \delta \in (0, 1]$ the δ-cut $C_\delta(\zeta(\cdot, \ldots, \cdot)) = C_\delta(\boldsymbol{x}^\star) := \{\boldsymbol{x} \in \mathbb{R}^n : \zeta(\boldsymbol{x}) \geq \delta\}$ is a closed compact and convex subset of $\mathbb{R}^n$.

One way to combine the characterizing functions $\xi_1(\cdot), \ldots, \xi_n(\cdot)$ into a vector characterizing function $\zeta(\cdot, \ldots, \cdot)$ of $\boldsymbol{x}^\star$ is the *minimum combination rule*

$$\zeta(x_1, \ldots, x_n) = \min[\xi_1(x_1), \ldots, \xi_n(x_n)] \quad \text{for all} \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

By this definition it holds that $\zeta(\cdot, \ldots, \cdot)$ is a vector characterizing function. Moreover, the δ-cuts $C_\delta(\boldsymbol{x}^\star)$ are Cartesian products of the δ-cuts of the n non-precise numbers $x_1^\star, \ldots, x_n^\star$, i.e.

$$C_\delta(\boldsymbol{x}^\star) = C_\delta(x_1^\star) \times C_\delta(x_2^\star) \times \cdots \times C_\delta(x_n^\star) \quad \text{for all} \quad \delta \in (0, 1].$$

Let us now consider a real valued continuous function $g(\cdot, \ldots, \cdot)$ which is applied to the non-precise numbers $x_1^\star, \ldots, x_n^\star$. The resulting value $g(x_1^\star, \ldots, x_n^\star)$ is again a non-precise number, denoted by $y^\star$. The values $\eta(y)$ of the characterizing function $\eta(\cdot)$ of $y^\star$ can be obtained by the *extension principle* developed by Zadeh (see Bandemer and Näther, 1992, or Dubois and Prade, 2000):

$$\eta(y) = \left\{ \begin{array}{lll} \sup\{\zeta(\boldsymbol{x}) : g(\boldsymbol{x}) = y\} & \text{if} & g^{-1}(\{y\}) \neq \emptyset \\ 0 & \text{if} & g^{-1}(\{y\}) = \emptyset \end{array} \right\} \quad \text{for all} \quad y \in \mathbb{R}$$

The function $\eta(\cdot)$ of $y^\star$ is indeed a characterizing function whose δ-cuts are given by

$$C_\delta(y^\star) = \left[\min_{\mathbf{x} \in C_\delta(\boldsymbol{x}^\star)} g(\boldsymbol{x}), \max_{\boldsymbol{x} \in C_\delta(\boldsymbol{x}^\star)} g(\boldsymbol{x}) \right] \quad \text{for all} \quad \delta \in (0, 1]$$

(see Viertl, 1996).

3 Testing of hypotheses

A hypothesis testing problem may be regarded as a decision problem where decisions have to be made about the truth of two propositions, the null hypothesis H_0 and the alternative H_1. For precise data, the decision rules are based on a sample $x_1, \ldots, x_n$ of an underlying random variable X whose distribution P_θ $(\theta \in \Theta)$ is at least partially unknown. The decision is usually depending on a test statistic $g(x_1, \ldots, x_n)$ which is a function of the observations $x_1, \ldots, x_n$. The randomness of the sample is expressed by the assumption that the data are generated by a random sample $X_1, \ldots, X_n$ of X according to the model P_θ. The decision is then based on the test statistic

$$T = g(X_1, \ldots, X_n)$$

which is evaluated for the sample, resulting in the value $t = g(x_1, \ldots, x_n)$.

Usually we consider two-decision testing problems where a hypothesis is rejected or not. In this case the space of possible values of the test statistic T is decomposed into a rejection region R and its complement $R^c = A$, the acceptance region. Depending on the hypotheses H_0 and H_1, the rejection region R takes one of the forms:

$$(a)\ T \leq t_l, \quad (b)\ T \geq t_u,$$

or (1)

$$(c)\ T \notin (t_a, t_b),$$

where t_l, t_u, or t_a and t_b are certain quantiles of the distribution of T such that under H_0 the error probabilities are

$$(a)\ P(T \leq t_l) = \alpha, \quad (b)\ P(T \geq t_u) = \alpha,$$

or (2)

$$(c)\ P(T \leq t_a) = P(T \geq t_b) = \alpha/2.$$

α is the probability of rejecting H_0 if H_0 is true, it is also called *significance level* of the test. The cases (a) and (b) represent one-sided tests, case (c) corresponds to a two-sided test. The hypothesis H_0 is rejected (and therefore H_1 is accepted) if the value $t = g(x_1, \ldots, x_n)$ falls into the rejection region R.

An equivalent testing procedure is to calculate the p-value which is defined for cases (a), (b), and (c) as

$$(a)\ p = P(T \leq t), \quad (b)\ p = P(T \geq t),$$

or (3)

$$(c)\ p = 2\min[P(T \leq t), P(T \geq t)].$$

If the p-value is less than α, then H_0 is rejected (at the significance level α), otherwise H_0 is not rejected.

One could also think of situations which give rise to the formulation of a three-decision testing problem:

- accept H_0 and reject H_1,
- reject H_0 and accept H_1,
- both H_0 and H_1 are neither accepted nor rejected.

The need for formulating a three-decision testing problem was already indicated by Neyman and Pearson (1933). A typical example is accepting a new treatment, rejecting it, or recommending it for further study. Here, we distinguish between acceptance region A, rejection region R, and region N with neither acceptance nor rejection.

4 The Fuzzy *p*-value

Let us consider the case of having fuzzy data $x_1^\star, \ldots, x_n^\star$ which are to be used for a statistical test. According to Section 2, the value $t^\star = g(x_1^\star, \ldots, x_n^\star)$ of a continuous test statistic becomes fuzzy, and the fuzziness of $t^\star$ is expressed by its characterizing function $\eta(\cdot)$. This implies that usual decision rules as described in Section 3 can no longer be applied.

The problem can be solved by using the concept of the p-value. Let $\mathrm{supp}(\eta(\cdot))$ denote the support of the characterizing function $\eta(\cdot)$ which is defined as $\mathrm{supp}(\eta(\cdot)) := \{x \in \mathbb{R} : \eta(x) > 0\}$. In applications the support of $\eta(\cdot)$ is usually finite.

Let us consider a one-sided test situation according to cases (a) and (b) of the previous section. As a first step we can define the p-value for non-precise numbers $t^\star$ of a test statistic with characterizing function $\eta(\cdot)$ as

$$(a)\;\; p = P(T \leq t = \max \operatorname{supp}(\eta(\cdot))) \quad (b)\;\; p = P(T \geq t = \min \operatorname{supp}(\eta(\cdot))) \tag{4}$$

as a *precise value*. A similar definition can be derived for a two-sided test. An illustration of this definition is given by the following example.

Example 4.1 *We use a test statistic T which is standard normally distributed. This is a very typical situation for many classical tests. Moreover, the characterizing function $\eta(\cdot)$ of the fuzzy values $t^\star$ of the test statistic is assumed to be a symmetric triangular function with center at 0.7. This rather theoretical example of a characterizing function was chosen only for illustrative purposes. For realistic characterizing functions obtained from fuzzy data compare Viertl (2002).*

Figure 1 shows the density function $f(x)$ and the characterizing function $\eta(\cdot)$. We want to test the hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, where θ is the unknown parameter and θ_0 a fixed value (here we have $\theta_0 = 0$). This is a one-sided test where definition (b) from above has to be applied, and the p-value is indicated by the shaded area in the plot. We can compare the resulting p-value with a significance level α (e.g. $\alpha = 0.05$) which has to be fixed in advance, and conclude that H_0 cannot be rejected at the significance level $\alpha = 0.05$.

The previous definition (4) of the p-value for fuzzy values of a test statistic has some shortcomings. It would be more logical to obtain a p-value which also becomes fuzzy because a fuzzy p-value would include much more information than a precise number. For this reason we improve the above definition and introduce the fuzzy p-value which is denoted by $p^\star$.

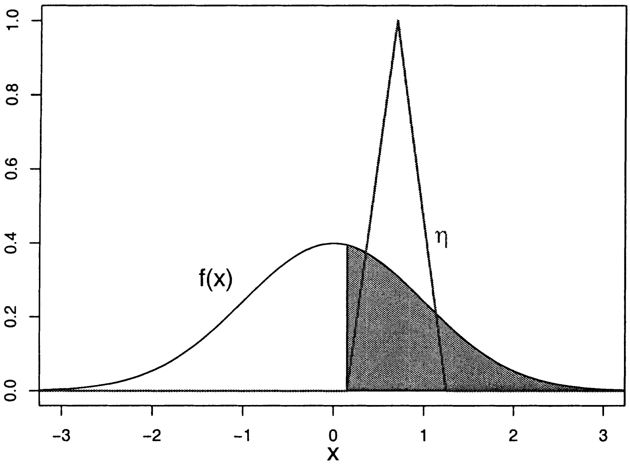

Fig. 1. One-sided test: Density $f(x)$ of a standard normally distributed test statistic and characterizing function $\eta(\cdot)$ of $t^\star$. The resulting p-value is indicated by the shaded area.

Since $\eta(\cdot)$ of $t^\star$ is a characterizing function, all δ-cuts ($\delta \in (0,1]$) are closed finite intervals $[t_1(\delta), t_2(\delta)]$. We can use these intervals for defining the corresponding intervals of fuzziness of $p^\star$. For a one-sided test we define according to cases (a) and (b) of Section 3

$$C_\delta(p^\star) = [P(T \leq t_1(\delta)), P(T \leq t_2(\delta))] \quad \text{for all} \quad \delta \in (0,1], \tag{5}$$

or

$$C_\delta(p^\star) = [P(T \geq t_2(\delta)), P(T \geq t_1(\delta))] \quad \text{for all} \quad \delta \in (0,1]. \tag{6}$$

In case of a two-sided test we first have to decide on which side of the median m of the distribution of the test statistic the most part of the amount of fuzziness of $t^\star$ is located. Therefore, we have to compute the area under the characterizing function $\eta(\cdot)$ of $t^\star$ which is on the left side of m and the area on the right side of m. We denote these areas by A_l and A_r, respectively, and define the intervals of fuzziness of $p^\star$ for a two-sided test by

$$C_\delta(p^\star) = \left\{ \begin{array}{ll} [2P(T \leq t_1(\delta)), \min[1, 2P(T \leq t_2(\delta))]] & \text{if } A_l > A_r \\ [2P(T \geq t_2(\delta)), \min[1, 2P(T \geq t_1(\delta))]] & \text{if } A_l \leq A_r \end{array} \right\} \text{ for all } \delta \in (0,1]. \tag{7}$$

Proposition 1: *The intervals $C_\delta(p^\star)$ of definitions* (5), (6), *and* (7) *are δ-cuts corresponding to a characterizing function $\xi(\cdot)$ of $p^\star$.*

Proof: We have to show that the three properties for characterizing functions are fulfilled. Properties (i) and (ii) follow immediately from the fact that $\eta(\cdot)$ of $t^\star$ is a characterizing function. Since the probabilities defining the intervals of the δ-cuts are restricted to $[0,1]$, we have for all $\delta \in (0,1]$ that the δ-cuts $C_\delta(p^\star)$ are closed finite intervals $[p_1(\delta), p_2(\delta)]$, which proves property (iii).

Note that $p_1(\delta) \geq 0$ and $p_2(\delta) \leq 1$ for all $\delta \in (0,1]$. Therefore, the δ-cuts of $p^\star$ can be interpreted in terms of probabilities and compared with the significance level α of the test. The decision is made according to a three-decision testing problem:

If, for all $\delta \in (0,1]$ and $p_1(\delta) \leq p_2(\delta)$,

- $p_2(\delta) < \alpha \longrightarrow$ reject H_0 and accept H_1
- $p_1(\delta) > \alpha \longrightarrow$ accept H_0 and reject H_1
- $\alpha \in [p_1(\delta), p_2(\delta)] \longrightarrow$ both H_0 and H_1 are neither accepted nor rejected.

In the third case, the uncertainty of making this decision is expressed by the characterizing function $\xi(\cdot)$ of $p^\star$.

The case $t_1(\delta) = t_2(\delta)$ for all $\delta \in (0,1]$ implies $p_1(\delta) = p_2(\delta)$. In this case we have a two-decision testing problem similar to tests with precise data.

5 Examples

In this section we want to illustrate different situations of statistical testing. In Examples 5.1 and 5.2 we use a normally distributed test statistic and consider a one-sided test. Example 5.3 is based on an F-distributed test statistic, the test is two-sided.

Example 5.1 *Similar to Example 4.1 we use a standard normally distributed test statistic T, and a symmetric triangular characterizing function $\eta(\cdot)$ of the fuzzy value $t^\star$ of the test statistic with center at 0.7. The left plot in Figure 2 shows the density function $f(x)$ and the characterizing function $\eta(\cdot)$.*

Like in Example 4.1 we want to test the hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, where θ is the unknown parameter and $\theta_0 = 0$ a fixed value. This is a one-sided test, and the fuzzy p-value is determined by definition (6). The resulting characterizing function $\xi(\cdot)$ of $p^\star$ is presented in the right plot of Figure 2.

In Figure 2 we show in detail how ξ was computed for $\delta = 0.5$. In the left plot, the δ-cut $C_\delta(t^\star) = [t_1(\delta), t_2(\delta)]$ for $\delta = 0.5$ is indicated by two vertical lines. The exact p-value for $t_2(0.5)$ is shown by the dark area, and for $t_1(0.5)$ by the dark and light area under the curve of $f(x)$. These p-values form the δ-cut $C_\delta(p^\star)$ for $\delta = 0.5$, which is presented in the right plot at the intersection of the horizontal line with the vertical lines through the exact p-values. Following this procedure for all $\delta \in (0, 1]$, the characterizing function $\xi(\cdot)$ of $p^\star$ can be constructed. Finally, we compare the resulting fuzzy p-value with a significance level α (e.g. $\alpha = 0.05$) which has to be fixed in advance, and conclude that H_0 is not rejected at the significance level $\alpha = 0.05$.

Example 5.2 *We take the same distribution of the test statistic and the same hypotheses as in Example 5.1. We also consider a symmetric, triangular-shaped characterizing function of $t^\star$ (see Figure 3). The aim here is to study fuzzy p-values for different outcomes of $t^\star$ which are presented by the characterizing functions η_1 to η_6. The corresponding characterizing functions ξ_1 to ξ_6 of $p^\star$ are shown in the right plot of Figure 3.*

We conclude that H_0 is rejected at the level $\alpha = 0.05$ for the first fuzzy t-value $t^\star$characterized by η_1 since for all $\delta \in (0, 1]$, the upper bound of the δ-cuts of ξ_1 is below 0.05. On the other hand, since the lower bound of the δ-cuts of ξ_3 to ξ_6 is above 0.05 for all $\delta \in (0, 1]$, H_0 cannot be rejected for values of $t^\star$ presented by η_3 to η_6. For an outcome of the test statistic $t^\star$ with characterizing function η_2 we can neither accept nor reject H_0 and H_1 at the significance level $\alpha = 0.05$.

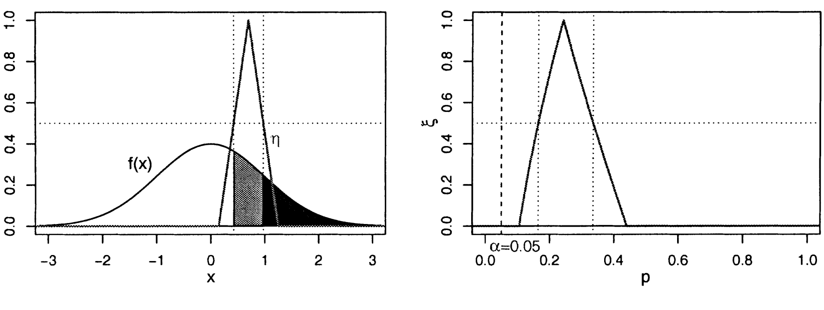

Fig. 2. One-sided test: Density $f(x)$ of a standard normally distributed test statistic and characterizing function $\eta(\cdot)$ of $t^\star$ (left), characterizing function $\xi(\cdot)$ of the fuzzy *p*-value $p^\star$ (right). The computation of the fuzzy *p*-value is indicated for $\delta = 0.5$. H_0 is not rejected at level $\alpha = 0.05$.

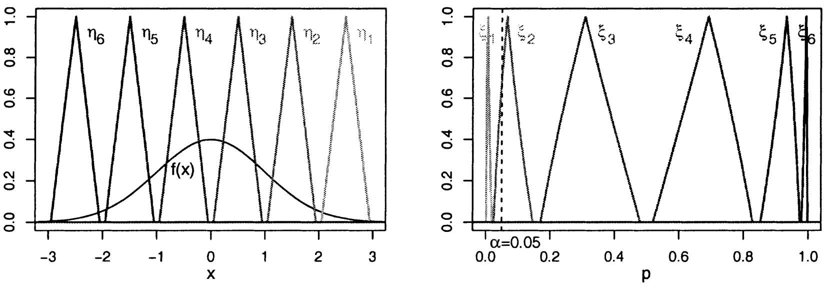

Fig. 3. One-sided test: Density $f(x)$ of a standard normally distributed test statistic and characterizing functions η_1 to η_6 for different outcomes of $t^\star$ (left); corresponding characterizing functions ξ_1 to ξ_6 of the fuzzy p-value (right). H_0 is rejected for η_1, not rejected for η_3 to η_6; no decision for η_2 (significance level $\alpha = 0.05$).

Example 5.3 *In this example we consider a two-sided test with the hypothesis* $H_0 : \theta = \theta_0$ *against* $H_1 : \theta \neq \theta_0$. *Moreover, we take a test statistic which is distributed according to* $F(10, 12)$, i.e. F*-distribution with 10 and 12 degrees of freedom. A typical example for such a situation is testing the equality of the variances of two random variables.*

Similar to the previous example we want to investigate different outcomes of the fuzzy value $t^\star$ *of the test statistic. These outcomes are described by the characterizing functions* η_1 *to* η_4 *and drawn in the left plot of Figure 4. The shape of the characterizing functions is similar to a normal distribution. We also present the density function* $f(x)$ *of the distribution* $F(10, 12)$*, and indicate its median m by a vertical line.*

In the right plot of Figure 4 we show the characterizing functions ξ_1 *to* ξ_4 *of* $p^\star$*, corresponding to* η_1 *to* η_4 *which are found by definition (7). For* η_1 *and* η_2 *the area* A_l *on the left side of the median m is zero, and for* η_4 *the area* A_r *is zero. For* η_3 *we see that* $A_l < A_r$ *and hence apply the second line of definition (7). Note that for computing* ξ_3 *we are bounding the* δ*-cuts of* $p^\star$ *at the maximum 1.*

Using the significance level $\alpha = 0.10$*, the hypothesis* H_0 *is rejected for test statistics* $t^\star$ *characterized by* η_1 *and* η_4*, and not rejected for* η_2 *and* η_3*.*

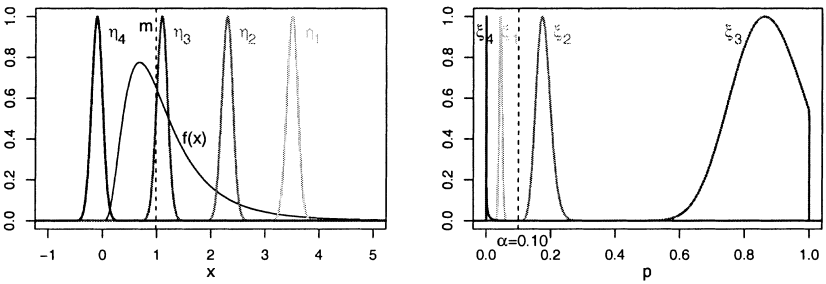

Fig. 4. Two-sided test: Density $f(x)$ of a test statistic with distribution $F(10, 12)$, median m, and characterizing functions η_1 to η_4 for different outcomes of $t^\star$ (left); corresponding characterizing functions ξ_1 to ξ_4 of the fuzzy p-value (right). H_0 is rejected for η_1 and η_4, and not rejected for η_2 and η_3 (significance level $\alpha = 0.10$).

Remark. *The situation with precise values t_0 of the test statistic is a special case of the methods given here by using the one-point indicator function $I_{\{t_0\}}(.)$ of the value t_0.*

6 Conclusions

Since observations of continuous random variables are non-precise also the values of related test statistics become non-precise. Therefore decision rules for statistical tests have to be adapted to this situation. This is done in the paper using a generalization of p-values. The resulting fuzzy p-values results in a more detailed description of the testing problem. Like for classical tests, the fuzzy p-value is compared with a given significance level α. The null hypothesis is then either rejected or not, or it comes to a third situation where no decision can be made, which is similar to sequential test decision procedures.

The situation of non-precise values of the test statistic shows the importance of a decision maker, because in real situations based on continuous quantities it is not possible to arrive at a decision by automatic decision rules like in standard statistics.

It should be possible to use the above concepts also in the situation of testing fuzzy hypotheses (see e.g. Last et al., 1999; Bertoluzza et al., 2002). This should be done in future research on testing based on fuzzy data.

References

[1] Afifi AA, Azen SP (1979) Statistical analysis: A computer oriented approach. 2nd edition, Academic Press, New York
[2] Bandemer H, Näther W (1992) Fuzzy data analysis. Kluwer, Dordrecht
[3] Bertoluzza C, Gil MA, Ralescu DA (eds) (2002) Statistical modeling, analysis and management of fuzzy data. Physica-Verlag, Heidelberg
[4] Dubois D, Prade H (eds) (2000) Fundamentals of fuzzy sets. Kluwer Acad. Publ., Boston
[5] Last M, Schenker A, Kandel A (1999) Applying fuzzy hypothesis testing to medical data. In Zhong N, Skowron A, Ohsuga S (eds) New directions in rough sets, data mining and granular-soft computing, Lecture Notes in Artificial Intelligence Series, Vol. 1711, Springer-Verlag, New York pp 221–229
[6] Lehmann EL (1986) Testing statistical hypotheses. 2nd edition, Wiley, New York
[7] Neyman J, Pearson ES (1933) The theory of statistical hypotheses in relation to probabilities a priori. Proc. Camb. Phil. Soc. 29:492–510
[8] Viertl R (1996) Statistical methods for non-precise data. CRC Press, Boca Raton, Florida
[9] Viertl R (2002) On the description and analysis of measurements of continuous quantities. Kybernetika 38(3):353–362