Control and Cybernetics

vol. **27** (1998) No. 3

Statistical inference about the median from vague data

by

Przemysław Grzegorzewski

Systems Research Institute, Polish Academy of Sciences,
ul. Newelska 6, 01-447 Warsaw, Poland
e-mail: pgrzeg@ibspan.waw.pl
and
Institute of Mathematics, Warsaw University of Technology,
Pl. Politechniki 1, 00-661 Warsaw, Poland

Abstract: In traditional statistics all parameters of the mathematical model and possible observations should be well defined. Sometimes such assumption appears too rigid for the real-life problems, especially when dealing with imprecise or linguistic data. To relax this rigidity fuzzy methods are incorporated into statistics. This paper is devoted to statistical inference about the population median in the presence of vague data. We propose the notion of fuzzy median. Then we suggest a fuzzy estimator and fuzzy confidence interval for the median. Next we discuss the problem of hypothesis testing concerning the median in the presence of imprecise data. All methods presented are distribution-free.

Keywords: confidence interval, estimation, fuzzy numbers, fuzzy sets, hypothesis testing, median, nonparametric statistics, vague data, the sign-test.

1. Introduction

In traditional statistics all parameters of the mathematical model and the observed experimental data should be well defined. However often this assumption appears too rigid for the real-life problems. We face such situations when our experimental data are imprecise or of linguistic type, like: "about five", "more or less seven", "not less then fifty", "approximately between seventeen and twenty", etc. Thus two types of uncertainty occur in our problem: randomness caused by a chance mechanism and vagueness brought about by the imprecise meaning of the data.

A possible way of handling situations like this is to apply the theory of fuzzy sets to describe vagueness and then generalize classical statistical methods to fuzzy data. The starting point is based on the pioneering papers of Zadeh on fuzzy sets, the extension principle and a concept of a linguistic variable Zadeh (1965, 1975). Next, Kwakernaak (1978, 1979) introduced the notion of a fuzzy random variable. Other definitions of fuzzy random variables are due to Kruse (1982, 1984), Puri and Ralescu (1986), Stein and Talati (1981). Kruse (1982) as well as Miyakoshi and Shimbo (1984) showed that the strong law of large numbers also holds for fuzzy random variables. Kruse (1984) suggested how to construct estimators under the presence of vagueness. Then Kruse and Meyer (1987, 1988) obtained fuzzy confidence intervals for the mean and the variance. They also proposed a method of testing statistical hypotheses for fuzzy data (Kruse, Meyer, 1987), which however has a lot of disadvantages (see Grzegorzewski, Hryniewicz, 1997). Attempts to contrive statistical problems with fuzzy data were also made by Casals, Gil and Gil (1986a, 1986b), Corral, Gil (1988), Gil (1988), Son, Song and Kim (1992) and Viertl (1996).

In all papers mentioned above the authors assume that a distribution in question is known except one parameter, e.g. the distribution is Gaussian with unknown mean and known standard deviation. Such an approach is called by statisticians parametric. Unfortunately we still don't have any effective goodness-of-fit test for fuzzy data. Thus we can not be sure that our fuzzy data have a distribution of a desired type indeed. Hence nonparametric methods would be useful in fuzzy statistics.

In this paper we show how to incorporate nonparametric methods into vague data problems. In Sec. 2 we define so called fuzzy median. Next we suggest how to estimate the median from vague data (Sec. 3) and how to construct a fuzzy confidence interval for the median (Sec. 4). Finally we discuss the problem of hypothesis testing concerning the median (Sec. 5). We generalize the well known sign-test into fuzzy sign-test. All methods presented are distribution-free, i.e. no assumptions on the type of the distribution are made.

2. Fuzzy random variables

The basic notion of the probability theory is a random variable. Roughly speaking, a random variable is a mapping which assigns to each random event a real number. A fuzzy random variable may be defined by analogy, however now we deal with fuzzy numbers. Thus we begin by recalling some basic concepts and notation connected with the notion of fuzzy number.

DEFINITION 2.1 *The fuzzy subset A of the real line $\mathbf{R}$, with the membership function $\mu : \mathbf{R} \to [0,1]$, is a fuzzy number iff*

(a) *A is normal, i.e. there exist an element $x_0 \in \mathbf{R}$ such that $\mu(x_0) = 1$;*

(b) *A is fuzzy convex, i.e. $\mu(\lambda x + (1-\lambda)y) \geq \mu(x) \wedge \mu(y)$ $\forall x, y \in \mathbf{R}$ and $\forall 0 \leq \lambda \leq 1$;*

(c) *μ is upper semicontinuous;*

(d) *supp (A) is bounded*

An useful tool for dealing with fuzzy numbers are their α-level sets. The α-level set A_α of a fuzzy number A is a nonfuzzy set defined as

$$A_\alpha = \{x \in R : \quad \mu(x) \geq \alpha\}.$$

The family $\{A_\alpha : \ \alpha \in [0,1]\}$ is a set representation of the fuzzy number A (see Kruse, Meyer, 1987). Basing on the resolution identity we have the alternative description of fuzzy numbers:

$$\mu(x) = \sup_{\alpha \in [0,1]} \{\alpha I_{A_\alpha}(x)\},$$

where $I_{A_\alpha}(x)$ denotes the indicator function of A_α.

Definition 2.1 implies that every α-level set of fuzzy number is a closed interval. Hence we have $A_\alpha = [A_\alpha^L, A_\alpha^U]$, where

$$\begin{aligned} A_\alpha^L &= \inf\{x \in \mathbf{R} : \mu(x) \geq \alpha\}, \\ A_\alpha^U &= \sup\{x \in \mathbf{R} : \mu(x) \geq \alpha\}. \end{aligned}$$

REMARK 2.1 *Some authors (see, e.g. Kruse, Meyer, 1987) consider separately the α-level sets and the strong α-cuts, i.e. sets of the form $\{x \in \mathbf{R} : \quad \mu(x) > \alpha\}$, $\alpha \in [0,1)$. This distinction, however, is useless in our case.*

A space of all fuzzy numbers will be denoted by $\mathbf{FN}(\mathbf{R})$. Of course, $\mathbf{FN}(\mathbf{R}) \subset \mathbf{F}(\mathbf{R})$, where $\mathbf{F}(\mathbf{R})$ is a space of all fuzzy sets on the real line.

Sometimes fuzzy sets are used to describe linguistic properties like: "rather less than 10", "greater than 50", etc. Such fuzzy sets are not fuzzy numbers, because their supports are not bounded. On account of importance of these fuzzy sets in applications, we introduce a family of the left-sided fuzzy numbers and the right-sided fuzzy numbers defined as follows (see Grzegorzewski, 1998):

DEFINITION 2.2 *The fuzzy subset A of the real line $\mathbf{R}$, with the membership function $\mu : \mathbf{R} \to [0,1]$, is the left-sided fuzzy number (right-sided fuzzy number) iff*

(a) *A is normal;*

(b) *A is fuzzy convex;*

(c) *μ is upper semicontinuous;*

(d) *supp (A) is bounded only from the left side (only from the right side).*

As before, we can use the alternative description based on α-level sets and the resolution identity. Definition 2.2 implies that every α-level set of the left-sided fuzzy number is an interval bounded from the left side, while the right-sided fuzzy number has α-level sets bounded from the right side. Families of all left-sided and right-sided fuzzy numbers will be denoted by $\mathbf{FN}_{LS}(\mathbf{R})$ and $\mathbf{FN}_{RS}(\mathbf{R})$, respectively (obviously, $\mathbf{FN}_{LS}(\mathbf{R})$, $\mathbf{FN}_{RS}(\mathbf{R}) \subset \mathbf{F}(\mathbf{R})$).

Now we will introduce the notion of fuzzy random variable. Our definition

Suppose a random experiment is described as usual by a probability space $(\Omega, \mathcal{F}, P)$, where Ω is the set of all possible outcomes of the experiment, $\mathcal{F}$ is a σ-algebra of subsets of Ω (the set of all possible events) and the function P, defined on $\mathcal{F}$, is a probability measure.

DEFINITION 2.3 *A mapping $X : \Omega \to \mathbf{FN}(\mathbf{R})$ is called a fuzzy random variable (f.r.v.) if it satisfies the following properties:*

1. *$\{X_\alpha(\omega) : \alpha \in [0,1]\}$ is a set representation of $X(\omega)$ for all $\omega \in \Omega$,*
2. *for each $\alpha \in [0,1]$ both X_α^L and X_α^U defined as*
$$X_\alpha^L \equiv X_\alpha^L(\omega) = \inf X_\alpha(\omega),$$
$$X_\alpha^U \equiv X_\alpha^U(\omega) = \sup X_\alpha(\omega),$$
are real-valued random variables on $(\Omega, \mathcal{F}, P)$.

Thus a fuzzy random variable X can be considered as a perception of an unknown usual random variable $V : \Omega \to \mathbf{R}$, called an original of X. Let χ denote a set of all possible originals of X. If only vague data are available, it is of course impossible to show which of the possible originals is the true one. Therefore we can define a fuzzy set of χ, with a membership function $\nu : \chi \to \mathbf{FN}(\mathbf{R})$ given as follows:

$$\nu(V) = \inf \left\{ \mu_{X(\omega)}(V(\omega)) : \omega \in \Omega \right\},$$

which corresponds to the grade of acceptability that a fixed random variable V is the original of the fuzzy random variable in question.

A random variable is characterized by its probability distribution. However often we are interested only in some parameters of the distribution. These parameters (e.g. measures of location or dispersion, descriptors of symmetry or shape) play a key role in mathematical statistics. They are useful particularly in statistics of vague data, where handling with probability distributions of fuzzy random variables is rather complicated. Let us consider a parameter $\theta = \theta(V)$ of random variable V. This parameter may be viewed as an image of a mapping which assigns to each random variable V with distribution P_θ the considered parameter θ. However if we deal with a fuzzy random variable we cannot observe our θ directly, but only its vague image. Using this reasoning together with Zadeh's extension principle, Kruse and Meyer (1987) introduced the notion of fuzzy parameter of fuzzy random variable, also called a fuzzy perception of the parameter θ. It is defined as follows

DEFINITION 2.4 *A fuzzy perception of a parameter θ is a fuzzy set $\Lambda(\theta)$ with a membership function*

$$\mu_{\Lambda(\theta)}(t) = \sup \left\{ \inf_{\omega \in \Omega} \mu_{X(\omega)}(V(\omega)) : V \in \chi, \theta(V) = t \right\}, \ t \in \mathbf{R},$$

This notion is well defined since if our random variable is crisp, i.e. $X = V$, we get $\Lambda(\theta) = \theta$.

In this paper we restrict our considerations to the central tendency parameters (i.e. the representative value for the population, also called the location parameter). Most often people are interested in the mean of a random variable. This notion was also generalized to the case of fuzzy random variables (see Kwakernaak, 1978). But there are distributions which have no mean (e.g. the Cauchy distribution). Thus from our nonparametric view the median would be a more suitable location parameter. Moreover, the median is less affected by extremal values of random variable. This is the reason why the median is so appreciated in statistics as a parameter robust to outliers. Here we propose a definition of a fuzzy median of fuzzy random variable X. Let us recall that $\gamma \in \mathbf{R}$ is the median of the random variable V if it satisfies following inequalities:

$$D_V(\gamma^-) \leq 0.5 \leq D_V(\gamma),$$

where D_V denotes the distribution function of V. Using Zadeh's extension principle we may generalize this notion to the fuzzy context.

DEFINITION 2.5 *A fuzzy median of a f.r.v. X is a fuzzy set Γ with a membership function defined as*

$$\mu_\Gamma(t) = \sup\left\{\inf_{\omega\in\Omega} \mu_{X(\omega)}(V(\omega)) : V \in \chi,\ D_V(t^-) \leq 0.5 \leq D_V(t)\right\},\ t \in \mathbf{R}.$$

Thus a fuzzy median may be regarded as a (fuzzy) perception of the unknown usual median. The following theorem is true:

THEOREM 2.1 *The fuzzy median Γ of a f.r.v. X is a fuzzy number with a set representation Γ_α of the form $\Gamma_\alpha = [\Gamma^L_\alpha, \Gamma^U_\alpha]$, where*

$$\Gamma^L_\alpha = \inf\left\{t \in \mathbf{R} :\ D_{X^L_\alpha}(t^-) \leq 0.5 \leq D_{X^L_\alpha}(t)\right\}$$

and

$$\Gamma^U_\alpha = \sup\left\{t \in \mathbf{R} :\ D_{X^U_\alpha}(t^-) \leq 0.5 \leq D_{X^U_\alpha}(t)\right\}.$$

Proof: If $\{\Gamma_\alpha\}$ is a set representation of the fuzzy median then

$$\Gamma_\alpha = \big\{t \in \mathbf{R} :\ \exists V \in \chi \text{ with } D_V(t^-) \leq 0.5 \leq D_V(t)$$
$$\text{such that } V(\omega) \in X_\alpha(\omega)\ \forall \omega \in \Omega\big\}.$$

By Definition 2.3 $X^L_\alpha(\omega),\ X^U_\alpha(\omega) \in X_\alpha(\omega)\ \forall\omega \in \Omega$ and $\forall\alpha \in [0,1]$. So the medians of the random variables $X^L_\alpha(\omega)$ and $X^U_\alpha(\omega)$ belong to Γ_α for all $\alpha \in [0,1]$.

Let $\alpha \in [0,1]$, and $V \in X_\alpha$. Assume that γ is a median of V. Since

$$X^L(\omega) \leq V(\omega) \leq X^U(\omega) \quad \forall\omega \in \Omega$$

we have

$$D_{X_\alpha^L}(t) \geq D_V(t) \geq D_{X_\alpha^U}(t) \text{ for all } t \in \mathbf{R}.$$

Therefore

$$\inf\left\{t \in \mathbf{R}: \ D_{X_\alpha^L}(t^-) \leq 0.5 \leq D_{X_\alpha^L}(t)\right\} \leq \gamma$$

$$\leq \sup\left\{t \in \mathbf{R}: \ D_{X_\alpha^U}(t^-) \leq 0.5 \leq D_{X_\alpha^U}(t)\right\}.$$

Since it holds for all $\alpha \in [0,1]$, we conclude that $\Gamma_\alpha = [\Gamma_\alpha^L, \Gamma_\alpha^U]$, where

$$\begin{aligned} \Gamma_\alpha^L &= \inf\left\{t \in \mathbf{R}: \ D_{X_\alpha^L}(t^-) \leq 0.5 \leq D_{X_\alpha^L}(t)\right\}, \\ \Gamma_\alpha^U &= \sup\left\{t \in \mathbf{R}: \ D_{X_\alpha^U}(t^-) \leq 0.5 \leq D_{X_\alpha^U}(t)\right\}, \end{aligned}$$

is the set representation of the fuzzy median Γ. By the normality of a fuzzy random variable, we have $\Gamma_\alpha \neq \emptyset \ \forall \alpha \in [0,1]$. Since f.r.v. is fuzzy-convex, we get $\Gamma_{\alpha_1} \subseteq \Gamma_{\alpha_2} \ \forall \alpha_1 > \alpha_2 \in [0,1]$. Thus we conclude that Γ is a fuzzy number, which proves the theorem. ∎

3. Fuzzy point estimation

Let $V_1, V_2, \ldots, V_n$ be a random sample which is the outcome of a random experiment. The problem of point estimation is to give a good guess for an unknown parameter of the underlying distribution. The best known point estimator of the median is, so called, the sample median defined as

$$\hat{\gamma} = \begin{cases} V_{[\frac{n}{2}]+1:n} & \text{if n is odd,} \\ \frac{1}{2}\left(V_{\frac{n}{2}:n} + V_{\frac{n}{2}+1:n}\right) & \text{if n is even,} \end{cases}$$

where $V_{1:n} \leq V_{2:n} \leq \ldots \leq V_{n:n}$ denote order statistics of the sample (i.e. the original sample after arrangement in the increasing order of magnitude) and where $[x]$ is the largest integer less than or equal to x. It is known that if the sample is drawn from the distribution with the uniquely determined median, then $\hat{\gamma}$ is a consistent estimator.

Now consider the situation that the results of our random experiment are not precise but vague. We describe them by a fuzzy random sample $X_1, X_2, \ldots, X_n$ which may be considered as a fuzzy perception of the random sample $V_1, V_2, \ldots$, V_n. A natural question arises: is it possible to estimate precisely an unknown median on the basis of these vague observations? The answer is negative, of course, because in the presence of randomness and fuzziness we can infer with the precision no better than the precision of the experiment outcomes. The best we may get is a fuzzy perception of our unknown parameter defined above.

Thus our task is to obtain a fuzzy point estimator of the fuzzy median, which may be viewed as a perception of the unknown parameter. Basing on the considerations given above we may define a fuzzy estimator of the median

DEFINITION 3.1 *A fuzzy sample median $\hat{\Gamma}$ from the fuzzy random sample X_1, $X_2, \dots, X_n$ is a fuzzy set with a membership function $\mu_{\hat{\Gamma}} : (\mathbf{FN}(\mathbf{R}))^n \to [0,1]$ given as follows*

$$\mu_{\hat{\Gamma}}(X_1, X_2, \dots, X_n)(t) = \sup \left\{ \alpha I_{[\hat{\Gamma}^L_\alpha, \hat{\Gamma}^U_\alpha]}(t) : \alpha \in [0,1] \right\} \text{ for } t \in \mathbf{R},$$

where

$$\hat{\Gamma}^L_\alpha = \hat{\Gamma}^L_\alpha(X_1, X_2, \dots, X_n) = \begin{cases} (X^L_\alpha)_{[\frac{n}{2}]+1:n} & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left((X^L_\alpha)_{\frac{n}{2}:n} + (X^L_\alpha)_{\frac{n}{2}+1:n}\right) & \text{if } n \text{ is even,} \end{cases}$$

and

$$\hat{\Gamma}^U_\alpha = \hat{\Gamma}^U_\alpha(X_1, X_2, \dots, X_n) = \begin{cases} (X^U_\alpha)_{[\frac{n}{2}]+1:n} & \text{if } n \text{ is odd,} \\ \frac{1}{2}\left((X^U_\alpha)_{\frac{n}{2}:n} + (X^U_\alpha)_{\frac{n}{2}+1:n}\right) & \text{if } n \text{ is even.} \end{cases}$$

Here $(X^L_\alpha)_{k:n}$ denotes the k-th order statistic of the sample $(X_1)^L_\alpha, \dots, (X_n)^L_\alpha$, while $(X^U_\alpha)_{k:n}$ is the k-th order statistic of the sample $(X_1)^U_\alpha, \dots, (X_n)^U_\alpha$.

REMARK 3.1 *If the observations are not vague but crisp, our fuzzy sample median becomes a traditional (crisp) sample median.*

The algebraic properties of the fuzzy sample median, in particular, fuzzy convexity and normality, are stated by the theorem.

THEOREM 3.1 *The fuzzy sample median $\hat{\Gamma}$ from the fuzzy random sample X_1, $X_2, \dots, X_n$ is a fuzzy number.*

Proof: Without loss of generality we assume that the size of our sample n is odd (otherwise the reasoning is analogous).

(i) Let take any $\alpha \in [0,1]$. Suppose that $\hat{\Gamma}^L_\alpha = (X_i)^L_\alpha$, where $i \in 1, 2, \dots, n$. Hence there exist at least $\left[\frac{n}{2}\right] + 1$ observations X_k such that $(X_k)^U_\alpha \geq (X_i)^L_\alpha$, $k \in 1, 2, \dots, n$. By Definition 3.1 we get $\hat{\Gamma}^U_\alpha = (X^U_\alpha)_{[\frac{n}{2}]+1:n} \geq (X_i)^L_\alpha = \hat{\Gamma}^L_\alpha$. Thus $\hat{\Gamma}_\alpha = [\hat{\Gamma}^L_\alpha, \hat{\Gamma}^U_\alpha] \neq \emptyset$ for all $\alpha \in [0,1]$.

(ii) Let take any two $\alpha_1, \alpha_2 \in [0,1]$ such that $\alpha_1 > \alpha_2$. Suppose that

$$\hat{\Gamma}^L_{\alpha_1} = (X_i)^L_{\alpha_1}; \quad \hat{\Gamma}^U_{\alpha_1} = (X_j)^U_{\alpha_1}; \quad \hat{\Gamma}^L_{\alpha_2} = (X_l)^L_{\alpha_2}; \quad \hat{\Gamma}^U_{\alpha_2} = (X_k)^U_{\alpha_2}.$$

There is no loss of generality in assuming that $l = i$. We have to consider three cases:

1. if $j = k$ then $\hat{\Gamma}_{\alpha_1} = [\hat{\Gamma}^L_{\alpha_1}, \hat{\Gamma}^U_{\alpha_1}] = [(X_i)^L_{\alpha_1}, (X_j)^U_{\alpha_1}] \subseteq [(X_i)^L_{\alpha_2}, (X_j)^U_{\alpha_2}] = [\hat{\Gamma}^L_{\alpha_2}, \hat{\Gamma}^U_{\alpha_2}] = \hat{\Gamma}_{\alpha_2}$, because both X_i and X_j are fuzzy convex;
2. if $j \neq k$ and $(X_j)^U_{\alpha_2} \leq (X_k)^U_{\alpha_2}$ then we have $\hat{\Gamma}_{\alpha_1} = [\hat{\Gamma}^L_{\alpha_1}, \hat{\Gamma}^U_{\alpha_1}] = [(X_i)^L_{\alpha_1}, (X_j)^U_{\alpha_1}] \subseteq [(X_i)^L_{\alpha_2}, (X_j)^U_{\alpha_2}] \subseteq [(X_i)^L_{\alpha_2}, (X_k)^U_{\alpha_2}] = [\hat{\Gamma}^L_{\alpha_2}, \hat{\Gamma}^U_{\alpha_2}] = \hat{\Gamma}_{\alpha_2}$;
3. if $j \neq k$ and $(X_j)^U_{\alpha_2} > (X_k)^U_{\alpha_2}$ then it should happen that $(X_j)^U_{\alpha_1} < (X_k)^U_{\alpha_1}$ so we get $\hat{\Gamma}_{\alpha_1} = [\hat{\Gamma}^L_{\alpha_1}, \hat{\Gamma}^U_{\alpha_1}] = [(X_i)^L_{\alpha_1}, (X_j)^U_{\alpha_1}] \subseteq [(X_i)^L_{\alpha_1}, (X_k)^U_{\alpha_1}] \subseteq$
$[(X_i)^L_{\alpha_2}, (X_k)^U_{\alpha_2}] = [\hat{\Gamma}^L_{\alpha_2}, \hat{\Gamma}^U_{\alpha_2}] = \hat{\Gamma}_{\alpha_2}$

Thus we have $\hat{\Gamma}_{\alpha_1} \subseteq \hat{\Gamma}_{\alpha_2}$ for any $\alpha_1 > \alpha_2$ and $l = i$. If $l \neq i$ the reasoning is analogous to that given above. Hence

- $\forall \alpha \in [0,1]$ $\hat{\Gamma}_\alpha$ is an interval (by the definition);
- $\forall \alpha \in [0,1]$ $\hat{\Gamma}_\alpha \neq \emptyset$;
- $\forall \alpha_1 > \alpha_2 \in [0,1]$ $\hat{\Gamma}_{\alpha_1} \subseteq \hat{\Gamma}_{\alpha_2}$

and we conclude that the fuzzy sample median $\hat{\Gamma}$ is convex. Moreover, by (i) $\hat{\Gamma}_{\alpha=1} \neq \emptyset$, so there exist such an element $x_0 \in \mathbf{R}$ that $\mu_{\hat{\Gamma}}(x_0) = 1$, and we see that $\hat{\Gamma}$ is normal. Since each α-level set of $\hat{\Gamma}$ is also bounded we conclude that the fuzzy sample median $\hat{\Gamma}$ is a fuzzy number, which completes the proof. ■

Now we will discuss statistical properties of the fuzzy sample median. We begin by recalling some basic concepts connected with the subject (see, e.g., Kruse, 1984, and Kruse, Meyer, 1987).

DEFINITION 3.2 *We say that a sequence $\{X_n\}_{n=1}^{\infty}$ of fuzzy random variables converges in probability to the fuzzy number Z (and we write $X_n \xrightarrow{p} Z$) if for every $\epsilon > 0$*

$$\sup_{\alpha \in [0,1]} P\left(\omega \in W:\ \left|(X_n(\omega))_\alpha^L - Z_\alpha^L\right| \vee \left|(X_n(\omega))_\alpha^U - Z_\alpha^U\right| > \epsilon\right) \to 0$$

as $n \to \infty$.

It is easily seen that this is a generalization of the convergence in probability for usual random variables to the case of fuzzy random variables.

Suppose that the unknown parameter θ has to be estimated from the vague data $X_1, X_2, \ldots, X_n$. Any mapping $\hat{\theta}_n(X_1, X_2, \ldots, X_n)$ from $(\mathbf{FN}(\mathbf{R}))^n$ into $\mathbf{FN}(\mathbf{R})$ may be considered as a fuzzy point estimator of that unknown parameter. However we need some criteria to choose a reasonable estimator among all possible ones. In the classical statistics such a basic property that a reasonable estimator should possess is consistency. In our case of vague data we may express this property in the following way.

DEFINITION 3.3 *Let $X_1, X_2, \ldots, X_n$ denote a fuzzy random sample from the distribution with unknown parameter θ and let $\Lambda(\theta)$ denote a fuzzy perception of θ based on $X_1, X_2, \ldots, X_n$ (see Definition 2.4). Then a fuzzy point estimator $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ is called a fuzzy-consistent estimator of the parameter θ (actually we have a sequence $\{\hat{\theta}_n\}_{n=1}^{\infty}$ of estimators) if for all sequences of fuzzy random variables $\{X_n\}_{n=1}^{\infty}$*

$$\hat{\theta}_n(X_1, X_2, \ldots, X_n) \xrightarrow{p} \Lambda(\theta).$$

Now we may prove a following theorem:

THEOREM 3.2 *The fuzzy sample median $\hat{\Gamma}$ is a fuzzy-consistent estimator of the*

Proof: Let θ denote the median we want to estimate (here we restrict our considerations only to distributions with unique median). If we have vague data $X_1, X_2, \ldots, X_n$, then the fuzzy median Γ is the fuzzy perception of γ, i.e. $\Lambda(\gamma) = \Gamma$. Let us consider the set representation $\{\Gamma_\alpha\}$ of the fuzzy sample median $\hat{\Gamma}$, i.e. $\hat{\Gamma}_\alpha = [\hat{\Gamma}_\alpha^L, \hat{\Gamma}_\alpha^U]$ for $\alpha \in [0,1]$. By Definition 3.1 it is easily seen that both $\hat{\Gamma}_\alpha^L = \hat{\Gamma}_\alpha^L(X_1, X_2, \ldots, X_n)$ and $\hat{\Gamma}_\alpha^U = \hat{\Gamma}_\alpha^U(X_1, X_2, \ldots, X_n)$ are usual estimators of the crisp medians γ_1 and γ_2 from crisp random samples $(X_1)_\alpha^L, (X_2)_\alpha^L, \ldots, (X_n)_\alpha^L$ and $(X_1)_\alpha^U, (X_2)_\alpha^U, \ldots, (X_n)_\alpha^U$, respectively. These medians are unique, because of the assumption on distributions (originals) under discussion. Since the usual sample median is a consistent estimator of the crisp median, provided that we restrict ourselves to distributions with unique median, we have $\hat{\Gamma}_\alpha^L \xrightarrow{p} \Gamma_\alpha^L$ and $\hat{\Gamma}_\alpha^U \xrightarrow{p} \Gamma_\alpha^U$ $\forall\alpha \in [0,1]$, where $\gamma_1 = \Gamma_\alpha^L$ and $\gamma_2 = \Gamma_\alpha^U$. Therefore $\forall\epsilon > 0$ and $\forall\alpha \in [0,1]$ we get $P\left(\omega \in W : \left|\hat{\Gamma}_\alpha^L(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) - \Gamma_\alpha^L\right| \vee \left|\hat{\Gamma}_\alpha^U(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) - \Gamma_\alpha^U\right| > \epsilon\right) \to 0$ as $n \to \infty$. Thus the fuzzy sample median $\hat{\Gamma}$ is a fuzzy-consistent estimator of the median, which completes the proof. ■

4. A fuzzy confidence interval for the median

Very often the experimenter is interested in finding an interval that contains the true (but unknown) parameter with a specified high probability. This is a problem of interval estimation. The desired interval is called the confidence interval and this specified probability is called the confidence level. Thus $\pi = [\pi_1, \pi_2]$, where π_1 and π_2 are functions of the observable random variables $V_1, V_2, \ldots, V_n$, is a confidence interval for the parameter θ on the confidence level $1 - \delta$ if

$$P\{\theta \in \pi\} \geq 1 - \delta.$$

Now we define a concept of the fuzzy confidence interval, due to Kruse and Meyer. Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and let denote by $\Lambda(\theta)$ the fuzzy perception of θ.

DEFINITION 4.1 *A fuzzy set Π is called a fuzzy confidence interval for θ on the confidence level $1 - \delta$ if*

$$\inf_{\alpha\in[0,1]} P\{\omega \in \Omega : \Lambda_\alpha \subseteq \Pi_\alpha\} \geq 1 - \delta,$$

where $\Pi_\alpha = [\Pi_\alpha^L, \Pi_\alpha^U]$ and $\Pi^L, \Pi^U : (\mathbf{FN}(\mathbf{R}))^n \to \mathbf{FN}(\mathbf{R})$.

Our definition is similar to those given in Kruse, Meyer (1987, 1988). If we know two usual (i.e. crisp) one-sided confidence intervals $[\pi_1, \infty)$ and $(-\infty, \pi_2]$ for θ we can also derive a fuzzy confidence interval for θ. This construction is also due to Kruse and Meyer (1987, 1988).

THEOREM 4.1 *Let $[\pi_1, \infty)$ and $(-\infty, \pi_2]$ be two usual one-sided confidence intervals for θ on the confidence level δ_1 and δ_2 respectively, where $\delta_1 + \delta_2 = \delta$, $\delta \in (0,1)$, and $\pi_1 \leq \pi_2$. Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and $\Lambda(\theta)$ denote a fuzzy perception of θ. Define for $\alpha \in [0,1]$*

$$\begin{aligned}
\Pi_\alpha^L(X_1, X_2, \ldots, X_n) &= \inf\{t \in \mathbf{R} : \forall i \in \{1, 2, \ldots, n\}\ \exists x_i \in (X_i)_\alpha \\
&\quad \textit{such that } \pi_1(x_1, x_2, \ldots, x_n) \leq t\}, \\
\Pi_\alpha^U(X_1, X_2, \ldots, X_n) &= \sup\{t \in \mathbf{R} : \forall i \in \{1, 2, \ldots, n\}\ \exists x_i \in (X_i)_\alpha \\
&\quad \textit{such that } \pi_2(x_1, x_2, \ldots, x_n) \geq t\}, \\
\mu_\Pi(t) &= \sup\{\alpha I_{[\Pi_\alpha^L, \Pi_\alpha^U]}(t) : \alpha \in [0,1]\}.
\end{aligned}$$

Then a fuzzy set Π with a membership function μ_Π is a fuzzy confidence interval for θ on the confidence level $1 - \delta$.

For the proof we refer the reader to Kruse, Meyer (1987). Basing on this theorem we may construct a confidence interval for the median.

THEOREM 4.2 *Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and $\delta \in (0,1)$. Define for $\alpha \in [0,1]$*

$$(\Pi_\Gamma(X_1, X_2, \ldots, X_n))_\alpha = \left[(X_\alpha^L)_{k_1+1:n}, (X_\alpha^U)_{k_2:n}\right],$$

where k_1 is chosen to be the largest integer which satisfies $\sum_{k=0}^{k_1} \binom{n}{k} (0.5)^n \leq \frac{\delta}{2}$ and $k_2 = n - k_1$. Then a fuzzy set Π_Γ with a membership function

$$\mu_{\Pi_\Gamma}(t) = \sup\left\{\alpha I_{\Pi_\Gamma(X_1, X_2, \ldots, X_n)}(t) : \alpha \in [0,1]\right\}$$

is a confidence interval for the median on the confidence level $1 - \delta$.

Proof: By Theorem 4.1 it suffices to show that for every fuzzy sample $X_1, X_2, \ldots, X_n$ and $\forall \alpha \in [0,1]$ $\left[\Pi_\alpha^L(X_1, X_2, \ldots, X_n), \Pi_\alpha^U(X_1, X_2, \ldots, X_n)\right] \subseteq (\Pi_\Gamma(X_1, X_2, \ldots, X_n))_\alpha$ is valid. Without loss of generality we assume that the sample size is odd.

It is known (e.g. see Gibbons, 1971) that if $V_1, V_2, \ldots, V_n$ denote a usual (i.e. crisp) random sample then the confidence interval for the median on the confidence level $1 - \delta$ has a form: $[V_{k_1+1:n}, V_{k_2:n}]$, where k_1 and k_2 are defined as in Theorem 4.2.

Let us set any $\alpha \in [0,1]$. Let us take $\xi \in \mathbf{R}$ such that there are $x_i \in (X_i)_\alpha$ $\forall i \in 1, 2, \ldots, n$ for which $\pi_1(x_1, x_2, \ldots, x_n) = x_{k_1+1:n} \leq \xi$ holds. Such ξ exists because the supports of X_i, $i = 1, 2, \ldots, n$, are finite. Since $\forall i \in \{1, 2, \ldots, n\}$ $x_i \geq (X_i)_\alpha^L$ is valid, it follows that

Therefore we get

$$\Pi_\alpha^L(X_1, X_2, \ldots, X_n) = \inf\{t \in \mathbf{R} : \forall i \in \{1,2,\ldots,n\}\ \exists x_i \in (X_i)_\alpha \text{ such that } \pi_1(x_1, x_2, \ldots, x_n) \le t\} \ge (X_\alpha^L)_{k_1+1:n}.$$

We can show in a similar way that

$$\Pi_\alpha^U(X_1, X_2, \ldots, X_n) = \sup\{t \in \mathbf{R} : \forall i \in \{1,2,\ldots,n\}\ \exists x_i \in (X_i)_\alpha \text{ such that } \pi_2(x_1, x_2, \ldots, x_n) \ge t\} \le (X_\alpha^U)_{k_2:n}.$$

These two inequalities show that $[\Pi_\alpha^L(X_1, X_2, \ldots, X_n), \Pi_\alpha^U(X_1, X_2, \ldots, X_n)] \subseteq [(X_\alpha^L)_{k_1+1:n}, (X_\alpha^U)_{k_2:n}] = (\Pi_\Gamma(X_1, X_2, \ldots, X_n))_\alpha$ which proves the assertion. ■

Sometimes one-sided confidence intervals are used in applications. Kruse and Meyer (1987) also showed how to derive one-sided fuzzy confidence intervals.

THEOREM 4.3 *Let $[\pi_1, \infty)$ and $(-\infty, \pi_2]$ be two usual one-sided confidence intervals for θ on the confidence level δ, $\delta \in (0,1)$. Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and $\Lambda(\theta)$ denote a fuzzy perception of θ.*

(i) *Define for $\alpha \in [0,1]$*

$$\begin{aligned} \underline{\Pi}_\alpha(X_1, X_2, \ldots, X_n) &= \inf\{t \in \mathbf{R} : \forall i \in \{1,2,\ldots,n\}\ \exists x_i \in (X_i)_\alpha \\ &\qquad \text{such that } \pi_1(x_1, x_2, \ldots, x_n) \le t\}, \\ \mu_{\underline{\Pi}}(t) &= \sup\left\{\alpha I_{[\underline{\Pi}_\alpha, \infty)}(t) : \alpha \in [0,1]\right\}. \end{aligned}$$

Then a fuzzy set $\underline{\Pi}$ with a membership function $\mu_{\underline{\Pi}}$ is the lower fuzzy confidence interval for θ on the confidence level $1-\delta$.

(ii) *Define for $\alpha \in [0,1]$*

$$\begin{aligned} \overline{\Pi}_\alpha(X_1, X_2, \ldots, X_n) &= \sup\{t \in \mathbf{R} : \forall i \in \{1,2,\ldots,n\}\ \exists x_i \in (X_i)_\alpha \\ &\qquad \text{such that } \pi_2(x_1, x_2, \ldots, x_n) \ge t\}, \\ \mu_{\overline{\Pi}}(t) &= \sup\left\{\alpha I_{(-\infty, \overline{\Pi}_\alpha]}(t) : \alpha \in [0,1]\right\}. \end{aligned}$$

Then a fuzzy set $\overline{\Pi}_\alpha$ with a membership function $\mu_{\overline{\Pi}}$ is the upper fuzzy confidence interval for θ on the confidence level $1-\delta$.

For more details we refer the reader again to Kruse, Meyer (1987). As in Theorem 4.2 we may construct the one-sided fuzzy confidence intervals for the median.

THEOREM 4.4 *Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and $\delta \in (0,1)$.*

(i) *Define for $\alpha \in [0,1]$*

$$(\underline{\Pi}_\Gamma(X_1, X_2, \ldots, X_n))_\alpha = \left[(X_\alpha^L)_{k_1+1:n}, \infty\right),$$

where k_1 is chosen to be the largest integer which satisfies

$$\sum^{k_1} \binom{n}{k} (0.5)^n \le \delta.$$

Then a fuzzy set $\underline{\Pi}_\Gamma$ *with a membership function*

$$\mu_{\underline{\Pi}_\Gamma}(t) = \sup\left\{\alpha I_{\underline{\Pi}_\Gamma(X_1,X_2,\ldots,X_n)}(t) : \alpha \in [0,1]\right\}$$

is the lower confidence interval for the median on the confidence level $1-\delta$.

(ii) *Define for* $\alpha \in [0,1]$

$$(\overline{\Pi}_\Gamma(X_1, X_2, \ldots, X_n))_\alpha = \left(-\infty, (X_\alpha^U)_{k_2:n}\right],$$

where k_2 *is chosen to be the smallest integer satisfying*

$$\sum_{k=k_2}^{n} \binom{n}{k} (0.5)^n \le \delta.$$

Then a fuzzy set $\overline{\Pi}_\Gamma$ *with a membership function*

$$\mu_{\overline{\Pi}_\Gamma}(t) = \sup\left\{\alpha I_{\overline{\Pi}_\Gamma(X_1,X_2,\ldots,X_n)}(t) : \alpha \in [0,1]\right\}$$

is the upper confidence interval for the median on the confidence level $1-\delta$.

The proof is similar to the proof of Theorem 4.2.

5. Fuzzy sign-test

In addition to estimation, one of the primary purposes of statistical inference is to test hypotheses. A statistical hypothesis is a statement about the population (or populations) from which one or more samples are drawn. The hypothesis under test is called the null hypothesis H_0. A statistical procedure which enables one to make a decision whether or not H_0 should be rejected is called a test. If the null hypothesis is rejected one accepts the alternative hypothesis H_1. A significance level δ is a preselected upper bound for a type I error, i.e. the error committed if the null hypothesis is rejected when it is true.

In the traditional approach to hypothesis testing all the concepts stated above are precise and well-defined and the theory of that problem has been explored thoroughly (see, e.g. Lehmann, 1986). However if we introduce vagueness into observations or hypotheses we face quite new and interesting problems. Diversity of approaches to testing hypotheses in fuzzy environment indicates that we are yet in the initial stage and the commonly accepted methodology has not been worked out. For a review of the achievements in this area we refer the reader to Grzegorzewski, Hryniewicz (1997). Here we present our view on the general problem of testing crisp hypothesis in the presence of vague data. Then we apply the submitted theory to distribution-free problems concerning the median.

Let $V_1, V_2, \ldots, V_n$ be a usual random sample from the population with unknown parameter θ_0. We consider the null hypothesis $H_0 : \theta = \theta_0$ against the two-sided alternative hypothesis $H_1 : \theta \neq \theta_0$ or against one of the following one-sided hypotheses $H_1' : \theta > \theta_0$ or $H_1'' : \theta < \theta_0$. To verify the null hypothesis on the significance level δ we use a test $\phi : \mathbf{R} \to 0, 1$ defined as follows

$$P(\phi(V_1, V_2, \ldots, V_n) = 1 | H_0) \le \delta$$

In statistics randomized tests $\phi : \mathbf{R} \to [0,1]$, which use an additional random mechanism, independent from a sample, are also known. Since their importance is rather of theoretical kind, here we restrict ourselves to non-randomized tests, called simply tests.

Now let us consider a fuzzy random sample $X_1, X_2, \ldots, X_n$. Grzegorzewski (1997) introduced the notion of fuzzy test for testing hypotheses in the presence of vague data.

DEFINITION 5.1 *Let $\delta \in (0,1)$ and let H_0 and H_1 denote the null hypothesis and the alternative hypothesis, respectively. A function $\varphi : (\mathbf{F}(\mathbf{R}))^n \to \mathbf{F}(\{0,1\})$ is called a fuzzy test for H_0 on the significance level δ if*

$$\sup_{\alpha \in [0,1]} P\left\{\omega \in \Omega : \varphi_\alpha\left(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\right) \subseteq \{1\} \,|H_0\right\} \leq \delta,$$

where φ_α is the α-level set of φ.

This definition reduces to the classical one if all observations are crisp.

It is well known that there is an equivalence between the totality of parameter values for which the null hypothesis is accepted and the structure of the confidence intervals (see, e.g. Lehmann, 1986). Thus having a confidence interval for a given parameter one may obtain easily a test for that parameter. Similarly, fuzzy confidence intervals can be used for the construction of fuzzy tests. That construction is due to Grzegorzewski (1997).

Let us denote by $\neg A$ the complement of a fuzzy set A, i.e. if μ_A is a membership function of A then a membership function of $\neg A$ is defined as $\mu_{\neg A}(x) = 1 - \mu_A(x), \forall x \in X$.

THEOREM 5.1 *Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and let $\delta \in (0,1)$. Let $\Pi = \Pi(X_1, X_2, \ldots, X_n)$ denote the two-sided fuzzy confidence interval for the parameter θ on the confidence level $1-\delta$. Then a function $\varphi : (\mathbf{F}(\mathbf{R}))^n \to \mathbf{F}(0,1)$ with its α-level sets defined as follows*

$$\varphi_\alpha(X_1, X_2, \ldots, X_n) = \begin{cases} \{0\} & \text{if} \;\; \theta_0 \in (\Pi_\alpha \setminus (\neg\Pi)_\alpha), \\ \{1\} & \text{if} \;\; \theta_0 \in ((\neg\Pi)_\alpha \setminus \Pi_\alpha), \\ \{0,1\} & \text{if} \;\; \theta_0 \in (\Pi_\alpha \cap (\neg\Pi)_\alpha), \\ \emptyset & \text{if} \;\; \theta_0 \notin (\Pi_\alpha \cup (\neg\Pi)_\alpha), \end{cases}$$

is a fuzzy test for the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ on the significance level δ.

By the theorem given above we may express a membership function of the fuzzy test considered above in a form more suitable for applications:

$$\begin{aligned} \mu_\varphi(x) &= \mu_\Pi(\theta_0) I_0(x) + \mu_{\neg\Pi}(\theta_0) I_1(x) = \\ &= \mu_\Pi(\theta_0) I_0(x) + (1 - \mu_\Pi(\theta_0)) I_1(x), \quad x \in \{0,1\} \end{aligned}$$

or, for short, $\varphi(x) = \mu_0/0 + (1-\mu_0)/1$, where $\mu_0 = \mu_\Pi(\theta_0)$. Thus it is seen that the fuzzy test, contrary to the classical crisp test, does not lead to the binary decision – to reject or to accept the null hypothesis – but to a fuzzy decision: we may get $\varphi = 0/0 + 1/1$ which indicates that we should reject H_0, or $\varphi = 1/0 + 0/1$ which means that H_0 should be accepted, but we may also get $\varphi(x) = \mu_0/0 + (1-\mu_0)/1$, where $\mu_0 \in (0,1)$, which may be interpreted as a degree of conviction that we should accept (μ_0) or reject ($1-\mu_0$) the hypothesis H_0.

It is worth noting that our fuzzy tests reduce do the usual (i.e. crisp) tests if the data are not vague but crisp.

We may also obtain fuzzy tests for one-sided hypotheses. In order to get a fuzzy test for testing hypothesis $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ it suffices to replace P in Theorem 5.1 by $\underline{\Pi}$. Similarly, to get a fuzzy test for testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, one has to replace P in Theorem 5.1 by $\overline{\Pi}$.

Now we can derive a fuzzy test for testing hypotheses concerning the median. This test is a natural generalization of the well-known sign-test (see, e.g., Gibbons, 1971) into situation with the presence of vague data.

THEOREM 5.2 *Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and let $\delta \in (0,1)$. Define for $\alpha \in [0,1]$*

$$\varphi_\alpha(X_1, X_2, \ldots, X_n$$
$$= \begin{cases} \{0\} & \text{if } ((X^L_\alpha)_{k_1+1:n} \vee (X^L_{1-\alpha})_{k_1+1:n}) \leq \gamma_0 \\ & \quad \leq ((X^U_\alpha)_{k_2:n} \wedge (X^U_{1-\alpha})_{k_2:n}), \\ \{1\} & \text{if } \gamma_0 < ((X^L_\alpha)_{k_1+1:n} \wedge (X^L_{1-\alpha})_{k_1+1:n}) \text{ or} \\ & \quad \gamma_0 > ((X^L_\alpha)_{k_1+1:n} \vee (X^L_{1-\alpha})_{k_2:n}) \\ \{0,1\} & \text{if } (X^L_\alpha)_{k_1+1:n} \leq \gamma_0 < (X^L_{1-\alpha})_{k_1+1:n} \text{ or} \\ & \quad (X^U_{1-\alpha})_{k_2:n} < \gamma_0 \leq (X^L_\alpha)_{k_1+1:n}), \\ \emptyset & \text{if } (X^L_{1-\alpha})_{k_1+1:n} \leq \gamma_0 < (X^L_\alpha)_{k_1+1:n} \text{ or} \\ & \quad (X^U_\alpha)_{k_2:n} < \gamma_0 \leq (X^L_{1-\alpha})_{k_1+1:n}, \end{cases}$$

where k_1 is chosen to be the largest integer which satisfies $\sum_{k=0}^{k_1} \binom{n}{k} (0.5)^n \leq \frac{\delta}{2}$ and $k_2 = n - k_1$. Then a function $\varphi : (\mathbf{F}(\mathbf{R}))^n \to \mathbf{F}(\{0,1\})$ with its α-level sets defined above is a test for the hypothesis that the median is equal to γ_0 against the alternative that it is not equal to γ_0 (i.e. $H_0 : \gamma = \gamma_0$ vs $H_1 : \gamma \neq \gamma_0$) on the significance level δ.

Proof: By Theorem 5.1 we know that a test for an unknown parameter with the two-sided alternative is completely determined by the confidence interval for that parameter and its complement. As it was shown in Theorem 4.2 that a fuzzy set $\Pi = \Pi(X_1, X_2, \ldots, X_n)$ with α-level sets defined as where k_1 is chosen to be the largest integer which satisfies $\sum_{k=0}^{k_1} \binom{n}{k} (0.5)^n \leq \frac{\delta}{2}$ and $k_2 = n - k_1$, is a two-sided confidence interval for the median γ_0 on the confidence level $1 - \delta$.

Our next claim is to find a complement of that confidence interval. A following lemma will be useful:

LEMMA 5.1

(a) *If $A \in \mathbf{F}_{LS}(\mathbf{R})$ then $\neg A \in \mathbf{F}_{RS}(\mathbf{R})$ and $(\neg A)_\alpha = \left(-\infty, A^L_{1-\alpha}\right]$.*

(b) *If $B \in \mathbf{F}_{RS}(\mathbf{R})$ then $\neg B \in \mathbf{F}_{LS}(\mathbf{R})$ and $(\neg B)_\alpha = \left[B^U_{1-\alpha}, \infty\right)$.*

The proof of the lemma is straightforward.

Since Π is a fuzzy number, it follows that

$$\Pi = \underline{\Pi} \cap \overline{\Pi},$$

where $\underline{\Pi}$ and $\overline{\Pi}$ are the left-sided and the right-sided fuzzy numbers respectively, with α-level sets defined as

$$\underline{\Pi}_\alpha = \left[(X^L_\alpha)_{k_1+1:n}, \infty\right) \text{ and } \overline{\Pi}_\alpha = \left(-\infty, (X^U_\alpha)_{k_2:n}\right].$$

Thus

$$\neg\Pi = \neg\underline{\Pi} \cup \neg\overline{\Pi}$$

and by the lemma given above we get

$$(\neg\Pi)_\alpha = \left(-\infty, (X^L_{1-\alpha})_{k_1+1:n}\right] \cup \left[(X^U_{1-\alpha})_{k_2:n}, \infty\right).$$

Thus a simple analysis lead as to the following conclusion

$$\begin{aligned}
&\overline{\Pi}_\alpha \setminus (\neg\overline{\Pi})_\alpha = \\
&= \left\{\gamma \in \mathbf{R} :\ ((X^L_\alpha)_{k_1+1:n} \vee (X^L_{1-\alpha})_{k_1+1:n}) \leq \gamma \leq ((X^U_\alpha)_{k_2:n} \wedge (X^U_{1-\alpha})_{k_2:n})\right\} \\
&(\neg\overline{\Pi})_\alpha \setminus \overline{\Pi}_\alpha = \left\{\gamma \in \mathbf{R} :\ \gamma < ((X^L_\alpha)_{k_1+1:n} \wedge (X^L_{1-\alpha})_{k_1+1:n}) \text{ or} \right. \\
&\left. \gamma_0 > ((X^L_\alpha)_{k_1+1:n} \vee (X^L_{1-\alpha})_{k_2:n})\right\} \\
&\overline{\Pi}_\alpha \cap (\neg\overline{\Pi})_\alpha = \left\{\gamma \in \mathbf{R} :\ (X^L_\alpha)_{k_1+1:n} \leq \gamma < (X^L_{1-\alpha})_{k_1+1:n} \text{ or} \right. \\
&\left. (X^U_{1-\alpha})_{k_2:n} < \gamma \leq (X^L_\alpha)_{k_1+1:n})\right\} \\
&\overline{\Pi}_\alpha \cup (\neg\overline{\Pi})_\alpha = \left\{\gamma \in \mathbf{R} :\ (X^L_{1-\alpha})_{k_1+1:n} \leq \gamma < (X^L_\alpha)_{k_1+1:n} \text{ or} \right. \\
&\left. (X^U_\alpha)_{k_2:n} < \gamma \leq (X^L_{1-\alpha})_{k_1+1:n}\right\}.
\end{aligned}$$

Hence

$$\varphi_\alpha(X_1, X_2, \ldots, X_n) = \begin{cases} \{0\} & \text{if } \gamma_0 \in (\overline{\Pi}_\alpha \setminus (\neg\overline{\Pi})_\alpha), \\ \{1\} & \text{if } \gamma_0 \in ((\neg\overline{\Pi})_\alpha \setminus \overline{\Pi}_\alpha), \\ \{0,1\} & \text{if } \gamma_0 \in (\overline{\Pi}_\alpha \cap (\neg\overline{\Pi})_\alpha), \\ \emptyset & \text{if } \gamma_0 \notin (\overline{\Pi}_\alpha \cup (\neg\overline{\Pi})_\alpha) \end{cases} =$$

$$= \begin{cases} \{0\} & \text{if } ((X_\alpha^L)_{k_1+1:n} \vee (X_{1-\alpha}^L)_{k_1+1:n}) \le \gamma_0 \le \\ & \quad ((X_\alpha^U)_{k_2:n} \wedge (X_{1-\alpha}^U)_{k_2:n}), \\ \{1\} & \text{if } \gamma_0 < ((X_\alpha^L)_{k_1+1:n} \wedge (X_{1-\alpha}^L)_{k_1+1:n}) \text{ or} \\ & \quad \gamma_0 > ((X_\alpha^L)_{k_1+1:n} \vee (X_{1-\alpha}^L)_{k_2:n}) \\ \{0,1\} & \text{if } (X_\alpha^L)_{k_1+1:n} \le \gamma_0 < (X_{1-\alpha}^L)_{k_1+1:n} \text{ or} \\ & \quad (X_{1-\alpha}^U)_{k_2:n} < \gamma_0 \le (X_\alpha^L)_{k_1+1:n}), \\ \emptyset & \text{if } (X_{1-\alpha}^L)_{k_1+1:n} \le \gamma_0 < (X_\alpha^L)_{k_1+1:n} \text{ or} \\ & \quad (X_\alpha^U)_{k_2:n} < \gamma_0 \le (X_{1-\alpha}^L)_{k_1+1:n}. \end{cases}$$

By Theorem 5.1 this completes the proof. ■

For one-sided alternatives we have

THEOREM 5.3 *Let $X_1, X_2, \ldots, X_n$ be a fuzzy sample and let $\delta \in (0,1)$.*

(i) *Define for $\alpha \in [0,1]$*

$$\varphi_\alpha(X_1, X_2, \ldots, X_n) = \begin{cases} \{0\} & \text{if } \gamma_0 \ge (X_\alpha^L)_{k_1+1:n} \vee (X_{1-\alpha}^L)_{k_1+1:n}, \\ \{1\} & \text{if } \gamma_0 < (X_\alpha^L)_{k_1+1:n} \wedge (X_{1-\alpha}^L)_{k_1+1:n}, \\ \{0,1\} & \text{if } (X_\alpha^L)_{k_1+1:n} \le \gamma_0 < (X_{1-\alpha}^L)_{k_1+1:n}, \\ \emptyset & \text{if } (X_{1-\alpha}^L)_{k_1+1:n} \le \gamma_0 < (X_\alpha^L)_{k_1+1:n}, \end{cases}$$

where k_1 is chosen to be the largest integer which satisfies $\sum_{k=0}^{k_1} \binom{n}{k} (0.5)^n \le \delta$. Then a function $\varphi : (\mathbf{F}(\mathbf{R}))^n \to \mathbf{F}(\{0,1\})$ with its α-level sets defined above is a fuzzy test for the hypothesis that the median is less or equal to γ_0 against the alternative that the true median exceeds the hypothesized value γ_0 (i.e. $H_0 : \gamma \le \gamma_0$ vs $H_1 : \gamma > \gamma_0$) on the significance level δ.

(ii) *A function $\varphi : (\mathbf{F}(\mathbf{R}))^n \to \mathbf{F}(\{0,1\})$ with its α-level sets defined as*

$$\varphi_\alpha(X_1, X_2, \ldots, X_n) = \begin{cases} \{0\} & \text{if } \gamma_0 \le (X_\alpha^U)_{k_2:n} \wedge (X_{1-\alpha}^U)_{k_2:n}, \\ \{1\} & \text{if } \gamma_0 > (X_{1-\alpha}^U)_{k_2:n} \vee (X_\alpha^U)_{k_2:n}, \\ \{0,1\} & \text{if } (X_{1-\alpha}^U)_{k_2:n} < \gamma_0 \le (X_\alpha^U)_{k_2:n}, \\ \emptyset & \text{if } (X_\alpha^U)_{k_2:n} < \gamma_0 \le (X_{1-\alpha}^U)_{k_2:n}, \end{cases}$$

where k_2 is chosen to be the smallest integer satisfying $\sum_{k=k_2}^{n} \binom{n}{k} (0.5)^n \le \delta$, is a fuzzy test for the hypothesis that the median is greater or equal to γ_0 against the alternative that it is less than γ_0 (i.e. $H_0 : \gamma \ge \gamma_0$ vs $H_1 : \gamma < \gamma_0$) on the significance level δ.

The proof is analogous to the previous one.

REMARK 5.1 *Theorem 5.3, part (i), remains valid if instead of fuzzy numbers we use the left-sided fuzzy numbers, while Theorem 5.3, part (ii) also holds if instead of fuzzy numbers we use the right-sided fuzzy numbers.*

6. Conclusions

It was shown how to estimate an unknown median, how to construct fuzzy the median in the presence of vague data, so frequent in the real life practice (environmetrics, medicine, social sciences, quality management, etc.). In fact, we have generalized techniques – well known in the traditional statistics – into more universal situation with fuzzy observations. These generalizations are natural, since if the data are precise, not vague, suggested procedures reduces to the traditional (i.e. crisp) ones.

The usefulness of the results stated above also lies in their distribution-free character. This is extremely important in the presence of vague data, because it is not known how to check the compatibility of a set of such observations with given distribution.

There are however, some open problems, e.g., how to verify fuzzy hypotheses, how to construct confidence intervals and statistical tests when the confidence level or significance level, respectively, is also not precise but vague, etc.

References

CASALS, R., GIL, M.A., GIL, P. (1986A) On the use of Zadeh's probabilistic definition for testing statistical hypotheses from fuzzy information. *Fuzzy Sets and Systems*, **20**, 175-190.

CASALS, R., GIL, M.A., GIL, P. (1986B) The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information. *European J. Oper. Res.*, **27**, 371-382.

CORRAL, N., GIL, M.A. (1988) A note on interval estimation with fuzzy data. *Fuzzy Sets and Systems*, **28**, 209-215.

GIL, M.A. (1988) Probabilistic – Possibilistic Approach to Some Statistical Problems with Fuzzy Experimental Observations, In: *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making*, Eds. J. Kacprzyk, M. Fedrizzi, Springer-Verlag, 286-306.

GIBBONS, J.D. (1971) *Nonparametric Statistical Inference.* McGraw-Hill, New York.

GRZEGORZEWSKI, P. (1997) Testing statistical hypotheses with vague data. *Fuzzy Sets and Systems*, to appear.

GRZEGORZEWSKI, P. (1998) Metrics and orders in space of fuzzy numbers. *Fuzzy Sets and Systems*, **97**, 83-94.

GRZEGORZEWSKI, P., HRYNIEWICZ, O. (1997) Testing hypotheses in fuzzy environment. *Mathware & Soft Computing*, **4**, 203-217.

KRUSE, R. (1982) The strong law of large numbers for fuzzy random variables. *Inform. Sci.*, **28**, 233-241.

KRUSE, R. (1984) Statistical estimation with linguistic data. *Inform. Sci.*, **33**, 197-207.

KRUSE, R., MEYER, K.D. (1987) *Statistics with Vague Data.* D. Riedel Publishing Company.

KRUSE, R., MEYER, K.D. (1988) Confidence Intervals for the Parameters of a Linguistic Random Variable. In: *Combining Fuzzy Imprecision with Prob-*

abilistic Uncertainty in Decision Making, Eds. J. Kacprzyk, M. Fedrizzi, Springer-Verlag, 113-123.

KWAKERNAAK, H. (1978, 1979) Fuzzy random variables, Part I: definitions and theorems. *Inform. Sci.*, **15**, 1-15; Part II: algorithms and examples for the discrete case. *Inform. Sci.*, **17**, 253-278.

LEHMANN, E.L. (1986) *Testing Statistical Hypotheses.* 2nd ed., Wiley, New York.

MIYAKOSHI, M., SHIMBO, M. (1984) A strong law of large numbers for fuzzy random variables. *Fuzzy Sets and Systems*, **12**, 133-142.

PURI, M.L., RALESCU, D.A. (1986) Fuzzy random variables. *J. Math. Anal. Appl.*, **114**, 409-422.

SON, J.CH., SONG, I., KIM, H.Y. (1992) A fuzzy decision problem based on the generalized Neyman-Pearson criterion. *Fuzzy Sets and Systems*, **47**, 65-75.

STEIN, W.E., TALATI, K. (1981) Convex fuzzy random variables. *Fuzzy Sets and Systems*, **6**, 271-283.

VIERTL, R. (1996) *Statistical Methods for Non-Precise Data.* CRC Press, Boca Raton, Florida.

ZADEH, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338-353.

ZADEH, L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning, Part I: *Inform. Sci.*, **8**, 199-249, Part II: *Inform. Sci.*, **8**, 301-375, Part III: *Inform. Sci.*, **9**, 43-80.