Mathware & Soft Computing 4 (1997) 203-217

Testing Statistical Hypotheses in Fuzzy Environment

Przemysław Grzegorzewski, Olgierd Hryniewicz
Systems Research Institute, Polish Academy of Sciences,
l ul. Newelska 6, 01-447 Warsaw, Poland,
e-mail: {pgrzeg, hryniewi}&ibspan.waw.pl

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 while dealing with linguistic data or imprecise requirements. To relax this rigidity fuzzy methods are incorporated into statistics. We review hitherto existing achievements in testing statistical hypotheses in fuzzy environment, point out their advantages or disadvantages and practical problems. We propose also a formalization of that decision problem and indicate the directions of further investigations in order to construct a more general theory.

Key words: hypothesis testing, fuzzy sets, fuzzy data, fuzzy hypothesis, the Neyman-Pearson lemma, the Bayesian approach.

1 Introduction

At the first stage of investigations concerning different kind of phenomena (e.g., biological, technical, physical, and social) initial hypotheses relating to these phenomena are often formulated. Then, during the second experimental stage, facts which either confirm or falsify these hypotheses are collected and analyzed. When the considered hypothesis is of a probabilistic nature, i.e., the phenomenon under investigation is described by a probabilistic model, the methods of mathematical statistics are used. These methods, called statistical tests, let us specify such events which are almost improbable when the considered hypothesis is true. Observations of such events indicate that the considered hypothesis may be not true, and therefore, should be rejected.

In traditional statistics all parameters of the mathematical model and the observed experimental data should be well defined. However, the complexity of investigated phenomena make the underlying models inadequate to the observed reality. In such situations the traditional models are valid only under some additional assumptions that might be not fulfilled. We face such a situation when our experimental data are of a linguistic type.

The problems stated above have motivated many researchers to enrich the traditional approach to hypothesis testing by introducing fuzzy models. These fuzzy models were proposed in order to describe uncertainties both in the experimental data and in the investigated hypotheses. Diversity of approaches indicates that we are yet in the initial stage, and the commonly accepted methodology hasn't been worked out. Therefore, it seems to be useful to summarize the already obtained results in order to show their weak points. This is the aim of Sections 4, 5, 6 and 7 of this paper. Such an analysis can also show the directions of further investigations in this area (Sec. 8). In Section 2 we review some of the basic notions of the traditional approach to hypotheses testing. Section 3 provides a formalization of the problem in order to make more visible the differences between the traditional and the new approaches.

2 Elements of the traditional theory of testing statistical hypotheses

Assume that the investigated phenomenon is described by a probability distribution P_θ which belongs to a family of distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. We consider the null hypothesis $H : \theta \in \Theta_H$ concerning the parameter θ, with the alternative hypothesis $K : \theta \in \Theta_K$, where Θ_H and Θ_K are subsets of Θ such that $\Theta_H \cup \Theta_K = \Theta$ and $\Theta_H \cap \Theta_K = \emptyset$. We assume that if θ were known one would also know whether or not the hypothesis is true. If Θ_H (Θ_K) contains only a single element, then the considered hypothesis is called "simple", otherwise is said to be "composite ".

In the experiment we observe a random variable $X = (X_1, X_2, \ldots, X_n)$, and this observation can lead to one of two possible decisions: either d_0 (to accept H, and to reject K), or d_1 (to accept K, and to reject H). Traditionally we denote d_0 by zero, and d_1 by one. Hence, a decision rule (a statistical test) can be defined as a function $\varphi : \mathcal{X} \to \{0, 1\}$, where $\mathcal{X}$ is a sample space. Such a test is also called a non-randomized statistical test, contrary to a randomized statistical test $\widetilde{\varphi} : \mathcal{X} \to [0, 1]$ which uses an additional random mechanism (independent from X) to undertake the decision. Each statistical test divides the sample space $\mathcal{X}$ into two exclusive subsets: $\{x \in \mathcal{X} : \varphi(x) = 0\}$ - the set of the acceptance of H, and $\mathcal{K} = \{x \in \mathcal{X} : \varphi(x) = 1\}$ - the set of the rejection of H (the acceptance of K) which is also called "a critical region". In practice, we compute a certain test statistic $T(X)$ (i.e. a function of the observations), then we find a critical region $\mathcal{K}$, and finally we reject the considered hypothesis if $T(X) \in \mathcal{K}$ or accept it otherwise.

Our decision can be either right or we can commit one of two possible errors: to reject H when it is true (type I error), or to accept H when it is false (type II error). The optimal test would minimize the probabilities of both errors simultaneously, but such a test doesn't exist. Therefore, in practice we set an upper limit for the probability of type I error: $E_\theta \varphi(x) \leq \delta$, $\forall \theta \in \Theta_H$ where δ is usually a small number (e.g. 0.01, or 0.05) and then we select a test which minimize the probability of type II error. In this case we also say that our test maximizes its power, i.e. the probability of the rejection of H when it is false. The power of test is of interest mainly while comparing different tests in order to find an optimal one. We say that the test φ is more powerful than the test ψ if

$$\begin{cases} E_\theta\varphi(x) \geq E_\theta\psi(x), & \forall\theta \in \Theta_K \\ E_\theta\varphi(x) > E_\theta\psi(x), & \text{for a certain } \theta \in \Theta_K. \end{cases}$$

The test which is more powerful than all other tests is called "uniformly most powerful".

The approach to the problem of statistical testing described above was established by Neyman and Pearson. There exist also other approaches, e.g. Bayesian and minimax. Let $L : \Theta \times \{d_0, d_1\} \to \mathcal{R}$ be a loss function. We interpret its value $L(\theta, d_i)$ as our loss when the true value of the parameter is θ and we undertake the decision d_i. The expected value of the loss function

$$R(\theta, \varphi) = E_\theta L(\theta, \varphi(x)), \qquad \theta \in \Theta_H$$

is called a risk.

In the Bayesian approach a prior probability distribution $\pi(\theta)$ is imposed on the set Θ. The prior distribution $\pi(\theta)$ describes initial subjective opinions (a prior information) about possible values of θ. Knowing the prior distribution we can calculate the Bayes risk

$$r(\pi, \varphi) = \int_\Theta R(\theta, \varphi)\pi(d\theta).$$

For a given prior distribution an optimal Bayes test φ is that one which minimizes the Bayes risk, i.e.

$$r(\pi, \varphi) \leq r(\pi, \phi), \qquad \forall\phi \in \Phi,$$

where Φ is a set of possible tests. Unfortunately, in order to apply this method it is necessary to assume that θ is a random variable and that its distribution $\pi(\theta)$ is known. This assumption is usually not warranted in applications. If no prior information regarding θ is available one might consider the maximum of the risk function to be the most important feature. Of two tests the one with a smaller maximum is then preferable, and the optimal test φ_0, called a minimax test, is that which minimizes the maximal risk,i.e.

$$\sup_{\theta\in\Theta} R(\theta, \varphi_0) = \inf_{\varphi\in\Phi} \sup_{\theta\in\Theta} R(\theta, \varphi).$$

Since this maximum represents the worst average loss that can result from the use of a given procedure, the minimax solution is one that gives the greatest protection against large losses. For more details concerning the traditional theory of testing statistical hypotheses we refer the reader to [7, 22].

3 A formal description of the hypotheses testing problem

The notions defined in the previous section can be used for a formal description of the hypotheses testing problem.

Definition 1 *Testing hypotheses is a decision problem described by an ordered 7-tuple*

$$(\mathcal{X}, \mathcal{P}, \mathcal{H}, \mathcal{D}, \Phi, \mathcal{U}, \mathcal{W}),$$

where: $\mathcal{X}$ is a sample space (possible states), $\mathcal{P}$ is a family of probability distributions, $\mathcal{H}$ is a set of hypotheses, $\mathcal{D}$ is a set of decisions, Φ is a set of tests, $\mathcal{U}$ is a utility function, and $\mathcal{W}$ is a preference function on the set of tests Φ.

The utility function $\mathcal{U}$ may be interpreted (depending on the applied approach) as a confidence level, the loss function or the risk, whereas the preference function $\mathcal{W}$ may be interpreted as the power of test, the Bayes risk, or the maximal risk. Both $\mathcal{U}$ and $\mathcal{W}$ we call sometimes requirements.

As it has been already stated in Introduction, the traditional approach might not be sufficient in the real life. A generalization of the classical theory that takes into account non-random uncertainty and lack of precision is really needed in practice. A fuzzy approach could be useful to describe all imprecision by "fuzzyfying" all elements of the 7-tuple which describes the decision problem in the presence of non-random uncertainties. However, it seems reasonable to restrict ourselves only to fuzzyfication of the space of possible states $\mathcal{X}$, and of the set of hypotheses $\mathcal{H}$. In the case of fuzzy states we can deal with imprecise observations such as: "is close to 5", "somewhere between 6 and 8", "a lot of", etc. In the case of fuzzy hypotheses we can formulate them, for example, in a form "$H : \theta$ is close to 5", etc. Fuzzy observations may lead to fuzzy decisions $\mathcal{D}$ and fuzzy tests Φ. The lack of precision can be also taken into account by the fuzzyfication of $\mathcal{U}$ or $\mathcal{W}$.

4 Testing hypotheses with fuzzy data

Testing hypotheses with fuzzy data is discussed by Casals, Gil, and Gil [3, 4, 5, 8]. The starting point of their papers comes from the paper of Tanaka,Okuda, and Asai [31], who introduced the notion of the fuzzy information system $\mathcal{A} = \{(A_i, \mu_i) : i \in I\}$, i.e., a set of fuzzy observations which fulfill additionally the orthogonality constraint:

$$\sum_{i \in I} \mu_i(x) = 1, \qquad \forall x \in S,$$

where $A_i \subset S \subseteq \mathcal{R}$, I is a set of indexes, and $\mu_i : S \to [0,1]$ is a membership function. So we have $X = \mathcal{A}^n$.

In their papers, Casals,Gil, and Gil used Zadeh's concept of the probability of fuzzy event

$$\widetilde{P}(A_1, \dots, A_n) = \int_{S^n} \mu_1(x_1) \dots \mu_n(x_n) dP(x_1, \dots, x_n),$$

where P is a probability distribution on S^n.

For the model given above, Casals, Gil, and Gil [3] proposed a new version of the Neyman-Pearson Lemma which supplies the uniformly most powerful test for simple hypotheses $H : \theta = \theta_0$, $K : \theta = \theta_1$

$$\varphi(X_1, \ldots, X_n) = \begin{cases} 1, & \text{when } \widetilde{P_1}(X_1, \ldots, X_n) > c_0 \widetilde{P_0}(X_1, \ldots, X_n) \\ 0, & \text{otherwise} \end{cases}$$

for a given significance level δ.

Casals, Gil, and Gil proposed also a method of constructing the optimal Bayesian and minimax tests. The approach proposed by Casals, Gil, and Gil has, however, some weak points. First, a finite set of possible observations has to be specified in advance. This precludes the usage of their test in the cases of observations that are only countable or even uncountable. Secondly, the orthogonality constraint has to be fulfilled. That condition is rather of a technical nature and is not natural for practical applications. Moreover, it precludes the existence of two observations which are described simultaneously by nondecreasing or nonincreasing membership function (such as, for example, "rather greater than 50"). In order to weaken the orthogonality constraint in the fuzzy information system Casals, Gil, and Gil [3] consider, so called, almost surely scale invariant tests. They conclude that given any set consisting of a finite number of fuzzy events we may construct a corresponding fuzzy information system leading to the same inference. More precisely, for any test φ based on fuzzy observations $\mathcal{O} = \{(O_i, \nu_i) : i = 1, 2, \ldots, m\}$ there exist a corresponding fuzzy information system $\mathcal{A} = \{(A_i, \mu_i) : i = 1, 2, \ldots, m\}$ and an optimal almost surely scale invariant test φ' based on $\mathcal{A}$, for which probability of type I and type II error coincides with the corresponding probabilities of φ. $\mathcal{O}$ and $\mathcal{A}$ are connected in the following way:

$$\nu_1(x_1) \ldots \nu_m(x_m) = C\mu_1(x_1) \ldots \mu_m(x_m)$$

for almost all $(x_1, \ldots x_m) \in S^m$, where C is a given positive constant.

Unfortunately the assertion that we can find a fuzzy information system $\mathcal{A}$ corresponding to any space of fuzzy observations $\mathcal{O}$ satisfying the equation given above is not true. A space $\mathcal{O}$ which includes at least two observations with nondecreasing or nonincreasing membership function is a simple counterexample. We can also prove a more general lemma:

Lemma 1 *Let S be a measurable space with a measure λ. Let $\mathcal{O} = \{(O_i, \nu_i) : i = 1, 2, \ldots, m\}$ denote a set of fuzzy observations, i.e., $O_i \subset S$ and $\nu_i : S \to [0, 1]$. If there exist at least two observations (O_j, ν_j) and (O_k, ν_k) in $\mathcal{O}$ such that:*

1. *$supp(\nu_j) \div supp(\nu_k) \supseteq U$, where $U \subset S$, $\lambda(U) > 0$ and $\div$ is a symmetric difference of sets;*

2. *$\nu_j(x) = \nu_k(x) = 1$ for $x \in V$, where $V \subset S$, $\lambda(V) > 0$;*

then it is not possible to construct the optimal almost surely scale invariant test corresponding to $\mathcal{O}$.

Proof. Without loss of generality we may assume that $m = 2$.

Let us take two fuzzy observations (O_1, ν_1), (O_2, ν_2) satisfying conditions 1), 2) of the lemma . Then we can take two points $a \in U$ and $b \in V$ such that

$$\begin{aligned} &\nu_1(a) = \alpha, \quad 0 < \alpha \leq 1 \\ &\nu_1(b) = 1, \\ &\nu_2(a) = 0, \\ &\nu_2(b) = 1. \end{aligned}$$

Suppose, on the contrary, that we can construct the optimal almost surely scale invariant test corresponding to any set of fuzzy observations $\mathcal{O}$. Then we should have a corresponding fuzzy information system (A_1, μ_1) and (A_2, μ_2) a positive constant C such that:

$$\nu_1(x_1)\nu_2(x_2) = C\mu_1(x_1)\mu_2(x_2) \qquad \text{for almost all } (x_1, x_2) \in S^2. \tag{1}$$

We thus get

$$\nu_1(a)\nu_2(a) = 0 = C\mu_1(a)\mu_2(a), \tag{2}$$

$$\nu_1(b)\nu_2(b) = 1 = C\mu_1(b)\mu_2(b), \tag{3}$$

$$\nu_1(a)\nu_2(b) = \alpha = C\mu_1(a)\mu_2(b), \tag{4}$$

$$\nu_1(b)\nu_2(a) = 0 = C\mu_1(b)\mu_2(a). \tag{5}$$

Since (A_1, μ_1) and (A_2, μ_2) satisfy the orthogonality constraint, we also have

$$\mu_1(a) + \mu_2(a) = 1, \tag{6}$$

$$\mu_1(b) + \mu_2(b) = 1. \tag{7}$$

From (2) we get $\mu_1(a) = 0$ or $\mu_2(a) = 0$, but from (4) we conclude that $\mu_1(a) \neq 0$ and $\mu_2(a) = 0$. Similarly, from (5) it follows that $\mu_1(b) = 0$ or $\mu_2(a) = 0$, but (3) gives $\mu_1(b) \neq 0$ and $\mu_2(a) = 0$. Hence (6) shows that $\mu_1(a) = 1$. Thus from (4) we get $\mu_2(b) = \alpha/C$. Therefore (3) leads us to the conclusion, that $\mu_1(b) = 1/\alpha$. If $\alpha < 1$ then $\mu_1(b) > 1$ which is impossible. If $\alpha = 1$ then $\mu_1(b) = 1$, but from (7) we get $\mu_2(b) = 0$ which is impossible as well.

The same reasoning applies to any choice of points $a \in U$ and $b \in V$. Thus equation (1) is not fulfilled on the set $U \times V$. Since $\lambda(U) > 0$ and $\lambda(V) > 0$ we have $\lambda(U \times V) > 0$. This contradicts the assertion that we can find a fuzzy information system leading to the almost surely scale invariant test corresponding to any set of observations and proves our lemma. ■

The reformulation of the Neyman-Pearson Lemma by Casals et al., though interesting, does not bring the problem close to reality. It is a well-known fact that the application of tests based directly on the Neyman-Pearson Lemma is minimal, and is limited only to the case of simple hypotheses. This lemma is extremely important as the starting point for the construction of other tests such as uniformly most powerful tests for composite hypotheses (the Karlin-Rubin theorem). It seems, however, that the generalization given by Casals, Gil, and Gil doesn't let us to make such an extension for testing composite hypotheses with fuzzy data. Moreover, the application of the test proposed by Casals et al. leads to very important computational problems. Practically these problems preclude the test from the usage even in the case of moderate samples. To overcome these problems Casals and Gil [5] proposed a modification based on the Central Limit Theorem. However, we don't know the accuracy of this approximation. Another problem arises when we analyze the conditions under which the test based on the Zadeh's definition of the probability could be used. In this approach the probability distribution is known explicitly, and the probabilities of events are not fuzzy. Hence, we deal with the case of uncertain observations of intrinsically non-fuzzy events driven by a probabilistic distribution of a known type, which raises some doubts. Moreover, no specific features of the fuzzy set theory are used except the notion of the membership functions, and thus this model could be regarded as a classical one only with a specific probability measure.

The problem of testing hypotheses with fuzzy data is also considered by Son, Song and Kim [29]. In many situations it is hard to find the uniformly most powerful test based on the Neyman-Pearson lemma or such a test does not even exist. To overcome these difficulties the locally most powerful test which maximizes the slope of the power function of the test at a certain desired point has been suggested. Son, Song and Kim present a reformulation of the generalized Neyman-Pearson lemma and propose a construction of the locally most powerful test for the fuzzy information system. Their paper has all weak points analogous to those discussed above. The reader may also have an impression that the considered problems could be solved on the basis of the traditional probability theory and the use of fuzzy sets is a little bit artificial.

5 Testing fuzzy hypotheses

The problem of testing fuzzy hypotheses has been considered independently by many authors. In their papers, techniques specific to fuzzy sets are widely used, contrary to the results described in the previous section. Therefore, new problems, non-existent in the traditional statistical approach, arise.

5.1 Bayesian approach

Delgado, Verdegay, and Vila [6] consider the Bayesian approach to the problem of testing fuzzy hypotheses. They use the Representation (Decomposition) Theorem and transform the original fuzzy hypothesis H to the problem of crisp hypotheses testing on a family of α-cuts

$$H_\alpha = \{\theta \in \Theta : \mu_H(\theta) > \alpha\}, \qquad \alpha \in [0, 1];$$

$$H = \bigcup_\alpha \alpha H_\alpha,$$

where $\mu_H : \Theta \to [0, 1]$ is a membership function of the hypothesis H.

As a result of this simple and natural transformation we arrive, however, at a fuzzy decision. To deal with this problem Delgado et al. proposed to consider the following quantity

$$\alpha_0 = \sup\{\alpha : H_\alpha \text{ is accepted}\}.$$

The value of α_0 describes a maximal α-cut for which the hypothesis is accepted. In their paper Delgado et al. considered two types of hypotheses. First, one-sided hypotheses defined by a membership function

$$\mu_H(\theta) = \begin{cases} 0, & \theta < \theta_0 \\ \Gamma(\theta), & \theta_0 \leq \theta < \theta_1 \\ 1, & \theta \geq \theta_1 \end{cases}$$

where $\Gamma : \mathcal{R} \to [0,1]$ is a continuous non decreasing (non increasing) function such that $\Gamma(\theta_0) = 0$, $\Gamma(\theta_1) = 1$ (or $\Gamma(\theta_0) = 1$, $\Gamma(\theta_1) = 0$). Secondly, two-sided hypotheses defined by a membership function

$$\mu_H(\theta) = \begin{cases} M\left(\frac{\theta_0-\theta}{a}\right), & \theta \leq \theta_0 \\ M\left(\frac{\theta-\theta_0}{a}\right), & \theta > \theta_0 \end{cases}$$

where $a \in \mathcal{R}^+$, $M : \mathcal{R} \to [0,1]$ is a non increasing smooth function such that $M(0) = 1$ and $\lim_{x\to\infty} M(x) = 0$.

The solution proposed by Delgado et al. has the virtue of simplicity. But it has its shortcomings: disregarding alternative hypotheses and considering only situation with a single observation.

Another approach was proposed by Saade and Schwarzlander [27]. They consider the problem of testing a crisp null hypothesis $H_0 : \theta = \theta_0$, where θ is the expected value of the Gaussian distribution, against a fuzzy alternative $H_1 : \theta = \theta_0 + a$, where a is a normal triangular fuzzy number. Similarly in Saade [26] a fuzzy null hypothesis $H_0 : \theta = \theta_0 + b$, where b is also a triangular fuzzy number, is considered. In both cases the authors arrive at decisions described by fuzzy numbers. To obtain a crisp decision they suggest to use either a "total distance criterion"

$$T(a) = \int_0^1 \frac{1}{2}\left(A'_\alpha + A''_\alpha\right) d\alpha$$

or "utility ranking index"

$$F_u(a) = \int_0^1 \ln\left(A'_\alpha A''_\alpha\right) d\alpha$$

where $A'_\alpha = \inf\{x \in \mathcal{R} : \mu_a(x) \geq \alpha\}$, $A''_\alpha = \sup\{x \in \mathcal{R} : \mu_a(x) \geq \alpha\}$, $\alpha \in [0,1]$.

In order to compare their test with a crisp minimax test Saade and Schwarzlander introduced as a preference function a "fuzzy Bayes risk" defined as follows:

$$FBR = \int_{Z_0} \left[P_0 c_{00} \lambda_{0R} + P_1 c_{01} \lambda_{1R}\right] dR + \int_{Z_1} \left[P_0 c_{10} \lambda_{0R} + P_1 c_{11} \lambda_{1R}\right] dR$$

where P_0 and P_1 are the prior probabilities of H_0 and H_1, respectively; c_{ij}, $i, j = 0$, describe the loss when the i-th hypothesis is accepted when the j-th is really true; Z_i, $i = 0, 1$, is a subset of the space of states for which the acceptance of the i-th hypothesis is preferred; λ_{iR} are so called fuzzy likelihood functions calculated for this test for given fuzzy numbers a and b (i.e. λ_{iR} is a fuzzy number induced from the fuzzy numbers a or b, respectively, by the map which is the conditional probability density function of the data under hypothesis H_i).

The results of Saade and Schwarzlander were obtained only for a specific case: under rather restrictive assumptions of one observation, triangular fuzzy numbers, and the Gaussian distribution. Moreover, the authors didn't analyze the choice of the appropriate index to compare two fuzzy numbers. However, their results are valuable for at least two reasons: both null and alternative hypotheses are considered and the obtained results were applied in practical situations [28].

Both preceding methods have all advantages and disadvantages of the Bayesian approach. Its weakest point is the necessity to know precisely a prior distribution of the considered parameter. But it is not our purpose to discuss here the principles of Bayesian analysis, which is a problem rather of a philosophical nature. We just want to point out only one strange feature of the methods presented above.

Both Delgado et al. and Saade et al. assume that the prior distribution does not depend on the level of α-cuts, so it is constant. This condition is, in our opinion, questionable. Because for crisp observations and precisely defined prior distribution, using well-known statistical methods one would get well-defined and unique posterior distribution. Thus when we have precise prior information, testing fuzzy (i.e. imprecise) hypotheses is useless. In practice, it seems more natural to consider situations with fuzzy prior information, and then testing fuzzy hypotheses. Such approach to Bayes estimation and reliability analysis has been adapted by Hryniewicz [11, 14] and Nagata [23].

5.2 Classical approach

Classical, i.e. Neyman-Pearson, approach to the hypotheses testing was considered in the paper of Watanabe and Imaizumi [33]. They propose to transform a fuzzy problem into a set of crisp problems defined for α-cuts. For a given α-cut a classical test problem is solved. Then the results are aggregated in a form of a fuzzy decision R represented by a membership function

$$\begin{aligned}
\chi_R(0) &= \begin{cases} \sup\limits_{\{r_\delta(0)=0\}} \mu_H(\theta), & \text{if } \{\theta : r_\delta(\theta) = 0\} \neq \emptyset \\ 0, & \text{if } \{\theta : r_\delta(\theta) = 0\} = \emptyset \end{cases} \\
\chi_R(1) &= 1 - \chi_R(0)
\end{aligned}$$

where

$$r_\delta(\theta) = \begin{cases} 0, & \text{if } T(\theta) \notin \mathcal{K}(\delta, \theta) \\ 1, & \text{if } T(\theta) \in \mathcal{K}(\delta, \theta) \end{cases},$$

$T(\theta)$ is a test statistics, and $\mathcal{K}(\delta, \theta)$ is a critical region for a given significance level δ.

Watanabe and Imaizumi propose also another solution in which they use a fuzzy critical function Φ described by a membership function

$$\chi_{\Phi}(j) = \begin{cases} \sup\limits_{\{r_\delta(0)=0\}} \mu_H(\theta), & \text{if } \{\theta : r_\delta(\theta) = j\} \neq \emptyset \\ 0, & \text{if } \{\theta : r_\delta(\theta) = j\} = \emptyset \end{cases}, \quad \text{for } j = 0, 1.$$

To investigate properties of fuzzy tests they use a minimal power function defined as

$$L_\delta^*(\theta_0) = \inf_{\{\theta:\mu_H(\theta)\geq\alpha, \alpha\in(0,1]\}} P\{T(\theta) \in \mathcal{K}(\delta,\theta)\,|\theta_0\}.$$

They suggest also another method of investigations with the help of a generalized power function $E_\theta\Phi$ defined as follows:

$$(E_\theta\Phi)_\alpha = E_\theta\Phi_\alpha$$

for $\alpha \in (0,1]$, where $(\cdot)_\alpha$ stands for the α-cut and the right hand side is defined by the Aumann integral [2].

Watanabe and Imaizumi proposed a general model which covers many specific cases, and illustrated their theoretical results with very good examples. The only point of criticism is related to their approach to the alternative hypotheses. At first, they set rather restrictive assumptions on the form of the alternative hypothesis, i.e. it should belong to the set

$$\overline{H_0} \cap \left[\left\{ \bigcup_{\alpha\in[0,1]} \alpha I \left(\bigcup_{\{\theta:\mu_H(\theta)\geq\alpha\}} \Theta_{K,\theta_0} \right) \right\}\right]$$

where $\overline{H_0}$ is a complement of a null hypothesis, and Θ_{K,θ_0} is a set of alternatives to the null hypothesis $H_0 : \theta = \theta_0$. But actually the construction of the test and consequently, the result of testing, does not depend on the alternative hypothesis.

6 Testing fuzzy hypotheses using fuzzy data

This general case was considered in the book of Kruse and Meyer [19]. Practical applications of the methodology proposed by Kruse and Meyer can be found in the papers of Höppner [9] and Höppner and Wolff [10] devoted to control charts in statistical quality control. Kruse and Meyer use the notion of a fuzzy random variable introduced by Kwakernaak [20]. In this approach fuzzy observations come from a probability distribution with a fuzzy parameter (e.g. from the Gaussian distribution with the mean expressed as "close to one"). Kruse and Meyer consider also fuzzy hypotheses (both one-sided and two-sided) which are described by convex and normal fuzzy numbers. It is easy to notice that applying the Kruse and Meyer approach to composite hypotheses it is necessary to introduce some ordering on the set of possible hypotheses. It stems from the fact that fuzzy numbers are not linearly ordered. In their book Kruse and Meyer use the following order

$$\forall\alpha \in [0,1] \qquad \mu \leq \nu \iff (\inf \mu_\alpha \leq \inf \nu_\alpha \quad \text{and} \quad \sup \mu_\alpha \leq \sup \nu_\alpha)$$

where μ, ν are two fuzzy numbers and μ_α, ν_α are their α-cuts. They are, however, not consequent, and in the case of one-sided hypotheses they don't use this order.

Kruse and Meyer transform a testing problem to N crisp tests on the set of α-cuts. Next, they reject the null hypothesis if at least k out of N α-levels lead to rejection. Both k and the set of α-levels $\{\alpha_1, \ldots, \alpha_N\} \subseteq [0, 1)$ are chosen arbitrary in advance. Here is the formal description of the test function:

$$\varphi(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if } \sum \varphi_i(X_1, \ldots, X_n) \geq k \\ 0, & \text{otherwise} \end{cases}$$

where

$$\varphi_i(X_1, \ldots, X_n) = \begin{cases} 1, & \text{if } T_{\alpha_i}(X_1, \ldots, X_n) \in \mathcal{K}_{\alpha_i}(\delta) \\ 0, & \text{otherwise} \end{cases},$$

T_α, $\mathcal{K}_\alpha$ are the α-cuts of the test statistics T and the critical region $\mathcal{K}$ (for a given significance level δ), respectively.

The method proposed by Kruse and Meyer is very simple to apply in practice. It is, however, open to a serious criticism. First, the choice of α-cuts and the number k is arbitrary. Therefore, by choosing appropriate α-cuts we can accept or reject any hypothesis. Secondly, the connection between the original hypothesis and the final one seems to be rather loose. It is not known how to choose k, and how should it depend on N. Moreover, it is sometimes assumed (see Höppner and Wolff [10]) that the tests on α-cuts are mutually independent; and this - in general - is not true. Therefore, the model of Kruse and Meyer despite its nice mathematical form of presentation is not sufficient for dealing with practical cases.

7 Other approaches

There are also a few papers, dealing with particular practical problems, in which a necessity of testing hypotheses under the lack of precision reveals. These are papers devoted to acceptance sampling [12, 13, 15, 24, 30], statistical process control [9, 10, 18, 21, 25, 32] or reliability analysis [11, 14, 16, 17]. However interesting, they are not universal and in general, they might be treated as particular cases of the situations considered above. So we do not discuss them here. But we want to mention briefly new ideas that could be found there.

7.1 Testing hypotheses with fuzzy requirements

Ohta and Ichihashi [24] design a single-sampling inspection plans by attributes when the producer and consumer's risk are not exact but approximate, e.g. "about δ". In fact, their problem is: given probabilities δ and β (i.e. type I error and type II error, respectively), construct a test for the binomial parameter (i.e. find a sample size n and an acceptance value c). But contrary to the classical situation, probabilities δ and β are not real but fuzzy numbers now. Their idea to test hypotheses with fuzzy requirements was applied in further papers [15, 16, 17, 18, 30]. This approach was extended by Arnold [1]. He considered statistical tests with continuously distributed test statistic under fuzzy constraints on the probabilities

δ and β of the errors. He also presented how to determine a test maximizing, so called, degree of satisfaction, which is a function of the errors and a sample size.

7.2 Testing fuzzy hypotheses with fuzzy data and requirements

This, up till now, most general problem have been considered by Hryniewicz in [12] and [13]. Though actually devoted to acceptance sampling, these papers contain a statistical test for the binomial parameter. Namely, the problem is how to construct an optimal test for vague data and imprecise admissible probabilities of both errors (i.e. type I error and type II error). Unfortunately, even for the crisp situation, there is no explicit solution generally. But in classical problem there exist many approximate methods to solve it. By means of the extension principle, Hald's approximate solution has been "fuzzyfied" and used for our fuzzy problem. Thus we've got fuzzy critical region $\mathcal{K}$. Using fuzzy addition and multiplication, we've got also fuzzy test statistic $T(X)$. As we have mentioned in Section 2, it suffices to check, whether $T(X) \in \mathcal{K}$, to undertake a decision. But in our case, we have to compare actually two fuzzy numbers. As it is known, there are many different methods for ranking fuzzy numbers. In order to find the best method for our decision problem an extensive simulation experiment was performed [12]. The results of simulations demonstrate rather strictly that, in general, none of the method is the best one (which is not surprising). They also show which methods are superior to other and which are totally unacceptable for our problem.
But the most important conclusion is that the result of a test depends strongly on the ordering criterion. Thus it is worth pointing out that two necessary conditions should be fulfilled. Firstly, we should fit a method of ranking, to the nature of the problem. Secondly, this method should be chosen before a testing procedure is started. Otherwise, any hypothesis could be accepted or rejected for the same data, just by the appropriate selection of the ordering criterion.

8 Concluding remarks

From the overview of the existing papers on the problem of testing statistical hypotheses in fuzzy environment it is easily seen that the results obtained yet are not sufficient for the construction of a more general theory. In our view such a theory should have a hierarchical structure of the form presented in Fig.1.

Statistical tests applied for practical tasks should be regarded as particular cases of the general theory. At present this ultimate goal seems to be rather distant. However, it should be kept in mind while building any new methodology of testing statistical hypotheses in fuzzy environment.

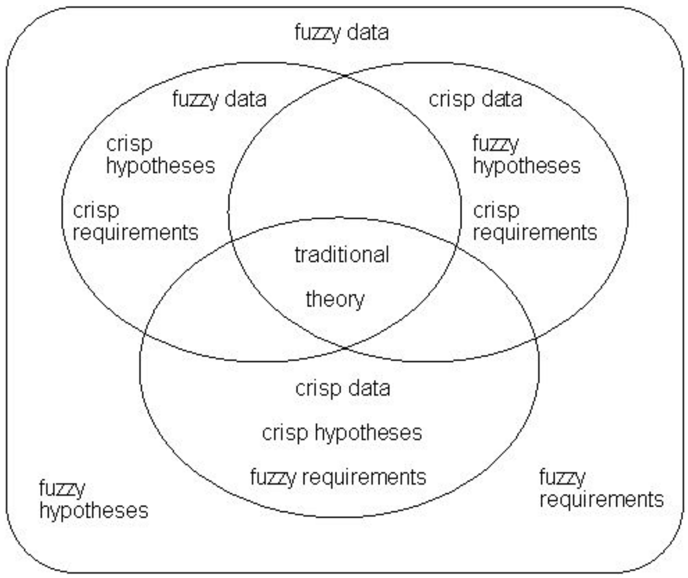

Fig. 1. The structure of a general theory of testing statistical hypotheses in fuzzy environment.

References

[1] Arnold B.F., Statistical Tests Optimally Meeting Certain Fuzzy Requirements on the Power Function and on the Sample Size, Fuzzy Sets and Systems 75 (1995), 365-372.

[2] Aumann R.J., Integrals of Set-Valued Function, J. Math. Anal. Appl. 12 (1965), 1-12.

[3] Casals R., Gil M.A., Gil P., On the Use of Zadeh's Probabilistic Definition for Testing Statistical Hypotheses from Fuzzy Information, Fuzzy Sets and Systems 20 (1986), 175-190.

[4] Casals R., Gil M.A., Gil P., The Fuzzy Decision Problem: an Approach to the Problem of Testing Statistical Hypotheses with Fuzzy Information, European J. Oper. Res. 27 (1986), 371-382.

[5] Casals R., Gil M.A., A Note on the Operativeness of Neyman-Pearson Tests with Fuzzy Information, Fuzzy Sets and Systems 30 (1989), 215-220.

[6] Delgado M., Verdegay J.L., Vila M.A., Testing Fuzzy Hypotheses. A Bayesian Approach, In: Approximate Reasoning In Expert Systems, Eds. M.M. Gupta,

A. Kandel, W. Bandler, J.B. Kiszka, Elsevier Science Publishers, 1985, 307-316.

[7] Ferguson T.S., Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York, 1967.

[8] Gil M.A., 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, 1988, 286-306.

[9] Höppner J., Statistiche Proceßkoontrolle mit Fuzzy-Daten, Ph.D. Dissertation, Ulm University, 1994.

[10] Höppner J., Wolff H., The Design of a Fuzzy-Shewart Control Chart, Research Report, Wrzburg University, 1995.

[11] Hryniewicz O., Estimation of Life-Time with Fuzzy Prior Information: Application in Reliability, In: Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making, Eds. J. Kacprzyk, M. Fedrizzi, Springer-Verlag, 1988, 307-321.

[12] Hryniewicz O., Acceptance Sampling for Imprecise Information from a Sample and Fuzzy Quality Criteria, Technical Report of SRI, Warsaw, 1992 (in Polish).

[13] Hryniewicz O., Statistical Decisions with Imprecise Data and Requirements,In: Systems Analysis and Decisions Support in Economics and Technology, Proceedings of the 9th Polish-Italian and 6th Polish-Finnish Conference, Eds. R. Kulikowski, K. Szkatua, J. Kacprzyk, Omnitech Press, 1994, 135–143.

[14] Hryniewicz O. , Lifetime Tests for Imprecise Data and Fuzzy Reliability Requirements, In: Reliability and Safety Analyses under Fuzziness, Eds. T. Onisawa, J. Kacprzyk, Physica-Verlag, Heidelberg, 1995, 169-179.

[15] Kanagawa A., Ohta H., A Design for Single Sampling Attribute Plan Based on Fuzzy Sets Theory, Fuzzy Sets and Systems 37 (1990), 173-181.

[16] Kanagawa A., Ohta H., Fuzzy Design for Fixed-Number Life Tests, IEEE Trans. Reliability 39 (1990), 394-398.

[17] Kanagawa A., Ohta H., Fixed- Time Life Tests Based on Fuzzy Life Characteristics, IEEE Trans. Reliability 41 (1992), 317- 320.

[18] Kanagawa A., Tamaki F., Ohta H., Control Charts for Process Average and Variability Based on Linguistic Data, Int. J. Prod. Res. 31 (1993), 913-922.

[19] Kruse R., Meyer K.D., Statistics with Vague Data, D. Riedel Publishing Company, 1987.

[20] Kwakernaak H., Fuzzy Random Variables, Part I: Definitions and Theorems, Inform. Sci. 15 (1978), 1-15; Part II: Algorithms and Examples for the Discrete Case, Inform. Sci. 17 (1979), 253-278.

[21] Laviolette M., Seaman J.W. jr., Barrett J.D., Woodall W.H., A Probabilistic and Statistical View of Fuzzy Methods, Technometrics 37 (1995), 249-261.

[22] Lehmann E.L., Testing Statistical Hypotheses, 2nd ed., Wiley, New York, 1986.

[23] Nagata Y., An Admissible Estimator in One-Parameter Exponential Family with Ambiguous Information, Ann. Inst. Stat. Math. 35 (1983), 193-199.

[24] Ohta H., Ichihashi H., Determination of Single-Sampling Attribute Plans Based on Membership Functions, Int. J. Prod. Res. 26 (1988), 1477-1485.

[25] Raz T., Wang J.H., Probabilistic and Membership Approaches in Construction of Control Charts for Linguistic Data, Production Plannaing & Control 1 (1990), 147-157.

[26] Saade J., Extension of Fuzzy Hypothesis Testing with Hybrid Data, Fuzzy Sets and Systems 63 (1994), 57-71.

[27] Saade J., Schwarzlander H., Fuzzy Hypothesis Testing with Hybrid Data, Fuzzy Sets and Systems 35 (1990), 197-212.

[28] Saade J., Schwarzlander H., Application of Fuzzy Hypothesis Testing to Signal Detection Under Uncertainty, Fuzzy Sets and Systems 62 (1994), 9-19.

[29] Son J.Ch., Song I., Kim H.Y., A Fuzzy Decision Problem Based on the Generalized Neyman-Pearson Criterion, Fuzzy Sets and Systems 47 (1992), 65-75.

[30] Tamaki F., Kanagawa A., Ohta H., A Fuzzy Design of Sampling Inspection Plans by Attributes, Japanese Journal of Fuzzy Theory and Systems 3 (1991), 315-327.

[31] Tanaka H., Okuda T., Asai K., Fuzzy Information and Decision in Statistical Model, In: Advances in Fuzzy Sets Theory and Applications, North-Holland, 1979, 303-320.

[32] Wang J.H., Raz T., On the Construction of Control Charts Using Linguistic Variables, Int. J. Prod. Res. 28 (1990), 477-487.

[33] Watanabe N., Imaizumi T., A Fuzzy Statistical Test of Fuzzy Hypotheses, Fuzzy Sets and Systems 53 (1993), 167-178.