Raport Badawczy

Research Report

RB/5/2002

Probabilistic Decision and Fuzzy Statistical Tests

O. Hryniewicz

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Olgierd Hryniewicz

Warszawa 2002

Possibilistic decisions and fuzzy statistical tests

Olgierd Hryniewicz

Systems Research Institute
Newelska 6, 01-447 Warsaw
POLAND
hryniewi@ibspan.waw.pl

Abstract

The paper deals with the problem of the interpretation of the results of statistical tests in terms of the theory of possibility. The well known in statistics concept of the observed test size p (also known as *p-value* and *significance*) has been given a new possibilistic interpretation and generalised for the case of imprecisely defined statistical hypotheses and vague statistical data. The proposed approach allows a practitioner to evaluate the test results using intuitive concepts of possibility, necessity or indifference.

Key words: Statistical decisions, Fuzzy statistical tests, Possibilistic indices.

1 Introduction

In the majority of practical cases decisions are made after the evaluation of existing information pertained to the considered problem. When this information is in a form of statistical data the decisions to be made are called statistical decisions. Let us assume that a phenomenon of interest is described by a random variable X (univariate or multivariate) distributed according to a certain probability distribution P_θ belonging to a family of probability distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ indexed by a parameter θ (one- or multidimensional). In a classical statistical setting we also assume that decisions depend entirely on the value of this parameter. If the value of θ were known we might take an appropriate decision without any problem. However, the value of θ is usually not known, and we could only formulate a respective hypothesis about it. Therefore, our decision is equivalent to the acceptance (or rejection) of a certain statistical hypothesis $H : \theta \in \Theta_H$ (where $\Theta_H \subset \Theta$) upon a value of the parameter θ. Such a hypothesis is called the null hypothesis. The

Preprint submitted to Elsevier Science *31 October 2002*

available statistical data either confirm or disprove our hypothesis. Statistical tests of this type were introduced by R.Fisher, and are known as statistical significance tests.

A better framework for making statistical decisions is offered by the Neyman-Pearson theory of statistical tests. According to this theory we should also formulate an alternative hypothesis $K : \theta \in \Theta_K$, where $\Theta_K \subset \Theta$, and $\Theta_H \cap \Theta_K = \emptyset$. In such a case we have either to reject H (and to accept K) or not to reject H (usually identified with the acceptance of H). To design the statistical test we usually set an upper value for the probability of a wrong rejection of the null hypothesis H (the so called probability of type I error). This probability, denoted by δ, is called *a significance level* of the test.

Despite their clear mathematical description, statistical tests are very often difficult to understand for a general user. First of all, the concept of significance level is not well understood, especially for single tests. Another reason of difficulties is the asymmetry between the null and the alternative hypotheses. It is clear, that only the probability of type I error δ is under control. The probability of wrong acceptance of the null hypothesis H (called the probability of type II error β) is generally either not explicitly defined or even not possible to define. Only certain statistical tests, such as procedures for testing *simple* statistical hypotheses, i.e. when $H : \theta = \theta_H$, and $K : \theta = \theta_K$, are relatively easy to explain for practitioners. However, even in this simple case there still exist some problems with practical interpretation of the results. Therefore, there is a need to provide the user with a methodology which let him/her to better understand the results of the statistical test. In the second section of the paper we propose to look at statistical tests from a perspective of the theory of possibility introduced by Zadeh [23]. We propose to use the possibilistic approach by Dubois and Prade [5] in order to arrive at new interpretation of the results of statistical tests.

In the classical theory of statistical tests all hypotheses should be well defined, and when the considered hypotheses are related to certain real life decisions their interpretation should be absolutely clear to potential users. Unfortunately, in many practical situations it is not the case. It is not difficult to indicate situations when precise formulation of statistical hypotheses is either useless or creates problems with clear understanding of the considered problem. To illustrate the problem let us consider an example taken from the statistical quality control when the decisions that are to be made depend upon a fraction of nonconforming (defective) items which have been found in a sample taken from a certain lot of products. Mathematically speaking the problem can be described as a certain statistical test for a value of a parameter p from the binomial (or hypergeometric) distribution. Statisticians formulate the problem as the test of the null hypothesis $H : \theta = p_0$ against the alternative $K : \theta = p_1$. In this setting the value of p_0 is called "an acceptable quality", and the value of p_1 is called "a disqualifying quality". However, in practice there is no ground to claim that these particular values have any special meaning. Their interpretation as fractions of nonconforming items in a lot for which that lot is accepted or rejected with a given probability is hardly understandable by practitioners who use rather vague concepts of "good" or "bad" quality. Therefore, for practical reasons it is much more convenient to formulate statistical hypotheses in a more relaxed fuzzy form. This is the motivation for the generalisation of classical statistical tests by fuzzy statistical tests with fuzzy requirements. The statistical tests with fuzzy requirements have been proposed by many authors such as Saade and Schwarzlander [19], Saade [18], Watanabe and Imaizumi [22], Arnold [2], Taheri and Behboodian [21], and Grzegorzewski and Hryniewicz [10]. In the third section of the paper we extend the possibilistic interpretation of statistical tests to the case of imprecisely defined fuzzy hypotheses.

Traditional statistical tests have been proposed for precisely defined crisp data. However, in many practical situations we face data which are not only random but vague as well. The introduction of vagueness to the problem of statistical testing leads to a new class of statistical tests which have been proposed by many authors such as Arnold [1], Casals et al. [3], Kruse and Meyer [13], Saade [18], Saade and Schwarzlander [19], Son et al. [20],Watanabe and Imaizumi [22], Römer and Kandel [17], and Montenegro et al. [16]. For deeper discussion and critical review of the problems considered there we refer the reader to the paper by Grzegorzewski and Hryniewicz [9]. Recently, Grzegorzewski [8] has proposed a unified approach for testing statistical hypotheses with vague data which is a direct generalisation of the classical approach. Unfortunately, all these proposals do not address the problem considered in this paper, namely the problem of the interpretation of the test result that is used for making a single and unique decision. In the fourth section of this paper we apply the approach proposed in the second section to the case of statistical tests with vague data. In the fifth section of the paper we combine the results presented in the previous sections in order to propose a unified possibilistic interpretation of fuzzy statistical tests when both statistical data and statistical hypotheses are given in a fuzzy form.

One of the most difficult problems that faces a practitioner is making the final decision basing on the interpretation of the results of a single statistical test. The reasons for this problem stem from the fact that the null and the alternative hypotheses are not symmetric. It means, that the result of the test depends upon which hypothesis is considered as the null hypothesis, and which as the alternative one. Thus, it may happen that on a given significance level δ we cannot neither reject the null hypothesis H vs. the alternative K, nor the alternative hypothesis K (treated as a new null hypothesis) vs. H (being a new alternative hypothesis). The problem becomes very serious when both considered hypotheses are "close", for example, when we test the null hypothesis $H : \theta \leq \theta_H$ against the alternative $K : \theta > \theta_H$. As the result of this asymmetry a decision maker is very often advised not to reject the null hypothesis despite the fact that the test data apparently support the alternative. This makes him confused, especially when the decision has to be made only once. To overcome this problem Hryniewicz [11] proposed a decision support tool that uses the notions of the theory of possibility. In the sixth section of this paper we present this approach as an additional tool that may be helpful for making final decisions.

2 Possibilistic interpretation of crisp statistical tests

The theory of statistical tests is well described in numerous textbooks. In the following paragraphs we recall some its most important notions. Let us observe a random sample $X_1, ..., X_n$, and the following decisions are to be made: either to reject H (and to accept K) or not to reject H (usually identified with the acceptance of H). Let's denote by 1 the rejection, and by 0, the acceptance of H. Hence, the decision rule, called a statistical test, can be defined as a function $\varphi : \mathcal{R}^n \rightarrow [0, 1]$. Each nonrandomised statistical test divides the whole space of possible observations of the random variable X into two exclusive subspaces: $\{(x_1, ..., x_n) \in \mathcal{R}^n : \varphi(x_1, ..., x_n) = 0\}$, and $\{(x_1, ..., x_n) \in \mathcal{R}^n : \varphi(x_1, ..., x_n) = 1\}$. First of these subspaces is called *an acceptance region*, and the second is called *a critical region*. In the majority of practical cases we deal with a certain *test statistic* $T = T(X_1, ..., X_n)$, and we reject the considered null hypothesis H when the value of T belongs to a certain critical region $\mathcal{K}$, i.e. if $T = T(X_1, ..., X_n) \in \mathcal{K}$. In such a case the decision rule looks like this

$$\varphi(X_1, ..., X_n) = \begin{cases} 1 \text{ if } T(X_1, ..., X_n) \in \mathcal{K}, \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

To define the critical region we must set an upper value for the probability of a wrong rejection of the null hypothesis H (the so called probability of type I error). This probability, denoted by δ, is called *a significance level* of the test. Thus, we have

$$P(\varphi(X_1, , ..., X_n) = 1|H) \leq \delta. \tag{2}$$

In general, for a given sample number n there may exist many statistical tests which fulfil this condition. However, only some of them may have additional desirable properties, and only those are used in practice. For more detailed description of the problem we refer the reader to textbooks on mathematical statistics such as the excellent book of Lehmann [15].

There exists a strong relationship between statistical tests and statistical confidence intervals. Suppose that the null hypothesis is given as $H : \vartheta \leq \vartheta_H$, and we observe a random sample $(X_1, \ldots, X_n)$. The one-sided confidence interval on a confidence level $1-\delta$ is given by $[\pi_L(X_1, \ldots, X_n; 1-\delta), \infty)$, and is closely related to the test of $H : \vartheta \leq \vartheta_H$ on a significance level δ. We reject the null hypothesis on the significance level δ if the observed value of $\pi_L(X_1, \ldots, X_n; 1-\delta)$ is larger than ϑ_H, i.e. when the inequality $\vartheta_H < \pi_L(x_1, \ldots, x_n; 1-\delta)$ holds. Similarly, we reject the hypothesis $H : \vartheta \geq \vartheta_H$ on the significance level δ when the inequality $\vartheta_H > \pi_U(x_1, \ldots, x_n; 1-\delta)$ holds, where $\pi_U(x_1, \ldots, x_n; 1-\delta)$ is the observed value of the upper limit of the one-sided confidence interval $(-\infty, \pi_U(X_1, \ldots, X_n; 1-\delta)]$ on a confidence level $1-\delta$. When we test the hypothesis $H : \vartheta = \vartheta_H$ on the significance level δ we reject it if either $\vartheta_H < \pi_L(x_1, \ldots, x_n; 1-\delta/2)$ or $\vartheta_H > \pi_U(x_1, \ldots, x_n; 1-\delta/2)$ holds, where $\pi_L(x_1, \ldots, x_n; 1-\delta/2)$ is the observed in the sample value of the lower limit of the two-sided confidence interval $\pi_L(X_1, \ldots, X_n; 1-\delta/2)$ on a confidence level $1-\delta$. The observed value of its upper limit $\pi_U(X_1, \ldots, X_n; 1-\delta/2)$ is given by $\pi_U(x_1, \ldots, x_n; 1-\delta/2)$. Thus, when we test a hypothesis about the value of the parameter ϑ we find a respective confidence interval, and compare it to the hypothetical value.

Dubois et al. [7] considered the problem of the relationship between statistical confidence intervals and possibility distributions. Following their approach we claim that the family of two-sided confidence intervals

$$[\pi_L(x_1, \ldots, x_n; 1-\delta/2), \pi_U(x_1, \ldots, x_n; 1-\delta/2)], \delta \in (0,1) \tag{3}$$

forms the *possibility distribution* $\tilde{\vartheta}$ of the estimated value of the unknown parameter ϑ. Similar observation holds also for the families of one-sided confidence intervals. The α-cuts of the membership function $\mu(\vartheta)$ denoted by $\left[\mu_L^{(\alpha)}, \mu_U^{(\alpha)}\right]$ are equivalent to the respective observed confidence intervals on a confidence level $1-\alpha$. Denote by $\tilde{\vartheta}_L$ the possibility distribution whose α-cuts are formed by the set of one-sided confidence intervals $\pi_L^{(\alpha)} = [\pi_L(x_1, \ldots, x_n; 1-\alpha), \infty)$. Thus, the membership function of $\tilde{\vartheta}_L$ is expressed as

$$\mu_L(\vartheta) = \sup\{\alpha I_{\pi_L^{(\alpha)}}(\vartheta) : \alpha \in [0,1]\}, \tag{4}$$

where $I_{\pi_L^{(\alpha)}}(\vartheta)$ denotes the characteristic function of $\pi_L^{(\alpha)}$. We can now look at the problem of testing the hypothesis $H : \vartheta \leq \vartheta_H$ as on the problem of the evaluation of the relation $\tilde{\vartheta}_L \succ \vartheta_H$.To evaluate this relation we propose to use the concept of the *Necessity of Strict Dominance* index introduced by Dubois and Prade [5] defined as

$$NSD = Ness\,(A \succ B) = 1 - \sup_{x,y;x\leq y} \min\left\{\mu_A\,(x)\,,\mu_B\,(y)\right\} \tag{5}$$

and NSD represents a *necessity* that the set A strictly dominates the set B.

Direct application of (5) for the evaluation of the relation $\tilde{\vartheta}_L \succ \vartheta_H$ leads to the following result

$$Ness\left(\tilde{\vartheta}_L \succ \vartheta_H\right) = 1 - \min\left(\mu_L(\vartheta), I\left(\vartheta_H\right)\right) \tag{6}$$

where

$$I\left(\vartheta_H\right) = \begin{cases} 1 & \vartheta = \vartheta_H \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

Thus, $Ness\left(\tilde{\vartheta}_L \succ \vartheta_H\right)$ is equal to one minus the ordinate of the intersection of $\mu_L(\vartheta)$ and the vertical line that crosses the x-axis at ϑ_H. From the definition of $\mu_L(\vartheta)$ it is easy to notice that the value of $Ness\left(\tilde{\vartheta}_L \succ \vartheta_H\right)$ is equal to the minimal value of the significance level of the test of the hypothesis H : $\vartheta \leq \vartheta_H$ for which this hypothesis has to be rejected. Therefore, the value of $Ness\left(\tilde{\vartheta}_L \succ \vartheta_H\right)$ is equivalent to the well known in statistics notion of the observed test size p, known also as *p-value* or *significance.*

It is not difficult to show that similar results are obtained when we consider statistical tests of hypotheses $H : \vartheta \geq \vartheta_H$ and $H : \vartheta = \vartheta_H$. In both cases, the Necessity od Strict Dominance indices are equal to the respective *p-values* of the considered tests.

The concept of *p-value* is not accepted by many statisticians as it does have frequency interpretation only in the case of simple statistical hypotheses. In the case of continuous probability distributions it can be related to the so called fiducial distributions of probability distribution parameters introduced by R.Fisher in the early 1930's. For a composite null hypothesis it can be computed as its fiducial probability. This approach, however, is also strongly criticised by many statisticians for the same reasons as the concept of *p-value.* The result presented above gives a new interpretation of the notion of *p-value* using concepts used in the theory of possibility.

3 Possibilistic interpretation of statistical tests of fuzzy hypotheses

In many practical situations both null and alternative hypotheses may not be formulated precisely. In such a case many authors propose to use the fuzzy sets theory, and to replace the sets θ_H (for the null hypothesis) and θ_K (for the alternative hypothesis) by their fuzzy equivalents $\widetilde{\theta}_H$, and $\widetilde{\theta}_K$, respectively. This approach allows a user to describe formally hypotheses that are defined imprecisely. For example, we can define a hypothesis in such vague terms as 'the average life time is about 5000 hrs.' or 'the fraction of defective items in a production lot is much smaller than 1%'. In all this cases, we may use fuzzy sets to represent imprecise notions like 'about 5000' or 'much smaller than 1'. The examples that illustrate this approach can be found in all fuzzy sets textbooks.

Let us assume that the null hypothesis H is defined as $H : \theta \in \widetilde{\theta}_H$, where $\widetilde{\theta}_H$ is a fuzzy subset of Θ described by a membership function $\mu_H(\theta)$. The membership function for the fuzzy set $\widetilde{\theta}_H$ we denote by $\mu_H(\theta)$.

Now, let us define the α-cuts of the fuzzy set $\widetilde{\theta}_H$ by

$$\theta_H^\alpha = \{\theta \in \Theta : \mu_H(\theta) \geq \alpha\}.$$

The set $\theta_H^\alpha = \left(\theta_{H,min}^\alpha, \theta_{H,max}^\alpha\right)$ is an ordinary set of real numbers, and for any number $\theta_H \in \theta_H^\alpha$ we may define a crisp null hypothesis $H : \theta \leq \theta_H$.

Proceeding as in the previous section we can treat the problem of testing the hypothesis $H : \theta \leq \widetilde{\theta}_H$ as the problem of the evaluation of the relation $\tilde{\vartheta}_L \succ \widetilde{\theta}_H$. From (5) we find that

$$Ness\left(\tilde{\vartheta}_L \succ \widetilde{\theta}_H\right) = 1 - \sup\min\left(\mu_L(\vartheta), \mu_H(\theta)\right). \tag{8}$$

Thus, $Ness\left(\tilde{\vartheta}_L \succ \widetilde{\theta}_H\right)$ is equal to one minus the ordinate of the intersection of $\mu_L(\vartheta)$ and the right-hand side of $\mu_H(\theta)$.

In the case of testing the statistical hypothesis $H : \theta \geq \widetilde{\theta}_H$ we have to evaluate the relation $\widetilde{\theta}_H \succ \tilde{\vartheta}_L$. In this case $Ness\left(\widetilde{\theta}_H \succ \tilde{\vartheta}_L\right)$ is equal to one minus the ordinate of the intersection of $\mu_U(\vartheta)$ and the left-hand side of $\mu_H(\theta)$. Similar results can be also obtained for the two-sided test of the hypothesis $H : \theta = \widetilde{\theta}_H$. In all these cases the obtained values of the NSD indices may be regarded as the generalisation of the observed test size p (*p-value* or *significance*) for the case of statistical tests of imprecisely defined hypotheses.

4 Possibilistic interpretation of statistical tests with fuzzy data

In the classical approach to statistical tests we assume that the observed random phenomenon is described by a crisp random variable X. Suppose now, that instead of a crisp random variable X we observe a fuzzy random variable $\widetilde{X}$. The notion of a fuzzy random variable has been defined by many authors. In this paper we use the definition proposed in Grzegorzewski[8].

Definition 1 (Grzegorzewski[8]). Let $(\Omega, \mathcal{A}, P)$ be a probability space, where Ω is a set of all possible outcomes of the random experiment, $\mathcal{A}$ is a σ-algebra of subsets of Ω (the set of all possible events), and P is a probability measure.

A mapping $\widetilde{X} : \Omega \to \mathcal{FN}(\mathcal{R})$, where $\mathcal{FN}(\mathcal{R})$ is the space of all fuzzy numbers, is called a fuzzy random variable if it satisfies the following properties:

(1) $\left\{\widetilde{X}_\alpha(\omega) : \alpha \in [0,1]\right\}$ is a set representation of $\widetilde{X}(\omega)$ for all $\omega \in \Omega$,
(2) for each $\alpha \in [0,1]$ both $\widetilde{X}^L_\alpha = \widetilde{X}^L_\alpha(\omega) = \inf \widetilde{X}_\alpha(\omega)$ and $\widetilde{X}^U_\alpha = \widetilde{X}^U_\alpha(\omega) = \sup \widetilde{X}_\alpha(\omega)$, are usual real-valued random variables on $(\Omega, \mathcal{A}, P)$.

This definition is similar to the definitions proposed by Kwakernaak [14] and Kruse [12], and the random variable $\widetilde{X}$ may be considered as a perception of an unknown usual random variable $X : \Omega \to \mathcal{R}$, called an original of $\widetilde{X}$.

Let $\widetilde{X}_1, \ldots, \widetilde{X}_n$ denote a fuzzy sample, i.e. a fuzzy perception of the usual random sample $X_1, \ldots, X_n$, from the population with the distribution P_Θ.

It is well known that in the statistical testing with crisp data there is an equivalence between the set of values of the considered probability distribution parameter for which the null hypothesis is accepted and a certain confidence interval for this parameter. The same equivalence exists in the case of statistical tests with fuzzy data.

Let us consider, for example, a statistical test with the null hypothesis $H : \vartheta \leq \vartheta_H$, and the alternative hypothesis $K : \theta > \theta_H$. In the case of crisp data there is one-to-one correspondence between the acceptance region for this test on the significance level δ and the one-sided confidence interval for the parameter θ on the confidence level $1 - \delta$. This correspondence has been described in the second section of this paper.

Kruse and Meyer[13] introduced the notion of a fuzzy confidence interval for the unknown parameter θ, when the data are fuzzy. In the considered case, a fuzzy equivalent of the lower limit of $[\pi_L(X_1, \ldots, X_n; 1-\delta), \infty)$ can be defined by the following α-cuts (for all $\alpha \in (0,1]$):

$$\begin{aligned} \Pi_L^{\alpha} &= \Pi_L^{\alpha}\left(\widetilde{X}_1,\ldots,\widetilde{X}_n;1-\delta\right) \\ &= \inf\left\{t \in \mathcal{R} : \forall i \in \{1,\ldots n\}\, \exists x_i \in \left(\widetilde{X}_i\right)_{\alpha} \right. \\ &\qquad \left. \text{such that } \pi_L\left(x_1,\ldots,x_n;1-\delta\right) \le t\right\} \end{aligned} \tag{9}$$

Similarly, we can define a fuzzy equivalent of the upper limit of the one-sided confidence interval $(-\infty, \pi_U(X_1,\ldots,X_n;1-\delta)]$ as given in Grzegorzewski [8]:

$$\begin{aligned} \Pi_U^{\alpha} &= \Pi_U^{\alpha}\left(\widetilde{X}_1,\ldots,\widetilde{X}_n;1-\delta\right) \\ &= \sup\left\{t \in \mathcal{R} : \forall i \in \{1,\ldots n\}\, \exists x_i \in \left(\widetilde{X}_i\right)_{\alpha} \right. \\ &\qquad \left. \text{such that } \pi_U\left(x_1,\ldots,x_n;1-\delta\right) \ge t\right\} \end{aligned} \tag{10}$$

where $\pi_U(x_1,\ldots,x_n;\delta) = \pi_L(x_1,\ldots,x_n;1-\delta)$. A similar definition can be also proposed for two-sided confidence intervals (see Kruse and Meyer[13] for details).

Now, let us construct the possibility distribution $\tilde{\vartheta}_F$ of the estimated value of the unknown parameter ϑ when sample data are fuzzy. First of all let us notice that in the case of fuzzy statistical data the limits of statistical confidence intervals become fuzzy for each confidence level $1-\delta$. In the second section of this paper we assumed that the value of δ (the significance level of the corresponding statistical test) is equal to the possibility degree α that defines the respective α-cut of the possibility distribution of $\tilde{\vartheta}$. We claim that in the possibilistic analysis of statistical tests on the significance level δ we should take into account only those possible values of the fuzzy sample whose possibility is not smaller than δ. Thus, the α-cuts of the membership function $\mu_F(\vartheta)$ denoted by $\left[\mu_{F,L}^{(\alpha)}, \mu_{F,U}^{(\alpha)}\right]$ are equivalent to the α-cuts of the respective fuzzy confidence intervals on a confidence level $1-\alpha$.

In the case of a statistical test with the null hypothesis $H : \theta \le \theta_H$ we denote the possibility distribution of the estimated value of the unknown parameter ϑ by $\tilde{\vartheta}_{F,L}$, and the respective α-cuts of its membership function $\mu_F(\vartheta)$, denoted by $\left[\mu_{F,L}^{(\alpha)}, \infty\right)$, are such that

$$\mu_{F,L}^{(\alpha)} = \inf_{\gamma \ge \alpha} \Pi_L^{\gamma}\left(\widetilde{X}_1,\ldots,\widetilde{X}_n;1-\delta\right) \tag{11}$$

The remaining part of the possibilistic analysis of the result of statistical test of the hypothesis $H : \theta \le \theta_H$ with fuzzy data is exactly the same as in the second section of this paper. We have to find the intersection point of the membership function $\mu_{F,L}(\vartheta)$ whose α-cuts are defined by (11) and the vertical line that crosses the x-axis at ϑ_H. One minus the ordinate of this point is equal to the NSD index of the relation $\tilde{\vartheta}_{F,L} \succ \vartheta_H$.

5 Possibilistic interpretation of statistical tests of fuzzy hypotheses in presence of fuzzy data

Let us consider the most general case when both statistical hypotheses and statistical data may be expressed in a vague form. To cope with this problem we have to combine the results from the previous sections. Let us assume that we observe a fuzzy random sample $\widetilde{X}_1, \ldots, \widetilde{X}_n$, and that we use this sample to test a fuzzy statistical hypothesis $H : \theta \in \widetilde{\Theta}_H$.

Suppose now that our fuzzy hypothesis is given as $H : \theta \leq \widetilde{\theta}_H$. In the case of crisp data we compare the lower limit of the one-sided confidence interval on a given confidence level $1 - \delta$ with the respective α-cut of the membership function that describes $\widetilde{\theta}_H$. In the possibilistic framework described in the third section it means that we compare the possibility distribution of $\tilde{\vartheta}_L$ with the fuzzy value of $\widetilde{\theta}_H$. In the presence of fuzzy data we have to compare the possibility distribution $\tilde{\vartheta}_{F,L}$ of the estimated value of the unknown parameter θ represented by its α-cuts given by (11) with the fuzzy value of $\widetilde{\theta}_H$. In such a case we have to find the intersection point of the membership function $\mu_{F,L}(\vartheta)$ and the left-hand side of $\mu_H(\theta)$. The NSD index of the relation $\tilde{\vartheta}_{F,L} \succ \tilde{\vartheta}_H$ is equal to one minus the ordinate of this point, i.e.

$$Ness\left(\tilde{\vartheta}_{F,L} \succ \widetilde{\theta}_H\right) = 1 - \sup \min\left(\mu_{F,L}(\vartheta), \mu_H(\theta)\right). \tag{12}$$

The NSD index defined by (12 can be regarded as the generalisation of the observed test size p (also known as *p-value* or *significance*) for the case of imprecisely defined statistical hypotheses and vague statistical data. In exactly the same way we can find the NSD index for other one-sided and two-sided statistical hypotheses.

6 Possibilistic interpretation of statistical decisions

As we have written in the Introduction one of the most difficult problems that faces a practitioner is to make decisions basing on the interpretation of the results of statistical tests. To overcome this problem Hryniewicz [11] proposed a new possibilistic interpretation of the results of statistical tests. In this interpretation he used the concept of the observed test size p (also known as *p-value* or *significance*). In the previous sections of this paper we have shown that the respective NSD indices can be interpreted as the generalised observed test size p. Thus, in the following part of this paper we denote by p the respective value of the NSD index.

Let us assume that our statistical decision problem is described, as usually, by setting two alternative hypotheses $H : \theta \in \Theta_H$ and $K : \theta \in \Theta_K$. Without the loss of generality we assume that $H : \theta \leq \theta_H$ and $K : \theta > \theta_K$, where $\theta_K \geq \theta_H$. According to Hryniewicz [11] we consider these two hypotheses *separately*. First, we the null hypothesis H. Using the methodology developed in the previous sections of this paper we could find the observed test size p_H for this hypothesis. The value of observed test size p_H shows how the observed data support the null hypothesis. When this value is relatively large we may say that the observed data strongly support H. Otherwise, we should say that the data do not sufficiently support H. It is worthwhile to note that in the latter case we do not claim that the data support K. The same can be done for the alternative hypothesis K, so we can find for this hypothesis the observed test size p_K. It is worthy to note that when $\Theta_H \cup \Theta_K = \Theta$ we have $p_K = 1 - p_H$.

Let us denote by 1 a situation when we decide that the data do not support the considered hypothesis, and by 0 a situation when we decide to accept the hypothesis. According to Hryniewicz [11] we propose to evaluate the null hypothesis H by a fuzzy subset $\widetilde{H}$ of $\{0, 1\}$ with the following membership function

$$\mu_{p_H}(x) = \begin{cases} \min[1, 2p_H] & \text{if } x = 0 \\ \min[1, 2(1 - p_H)] & \text{if } x = 1 \end{cases} \tag{13}$$

which may be interpreted as a *possibility distribution* of H. It is worthy to note that $\sup(\mu_{p_H}(0), \mu_{p_H}(1)) = 1$, and $\mu_{p_H}(1) = 1$ indicates that it is *plausible* that the hypothesis H is not true. On the other hand, when $\mu_{p_H}(0) = 1$ we wouldn't be surprised if H was true. It is necessary to stress here that the values of $\mu_{p_H}(x)$ do not represent the probabilities that H is false or true, but only a *possibility distribution* of the correctness of alternative decisions with respect to this hypothesis.

The same can be done for the alternative hypothesis K. We may evaluate the alternative hypothesis K by a fuzzy subset $\widetilde{K}$ of $\{0, 1\}$ with the following membership function

$$\mu_{p_K}(x) = \begin{cases} \min[1, 2p_K] & \text{if } x = 0 \\ \min[1, 2(1 - p_K)] & \text{if } x = 1 \end{cases} \tag{14}$$

which may be interpreted as a *possibility distribution* of the correctness of alternative decisions with respect to K.

To choose an appropriate decision, i.e. to choose either H or K Hryniewicz [11] proposes to use four measures of possibility defined by Dubois and Prade [5]. First measure is called the *Possibility of Dominance*, and for two fuzzy sets A and B is defined as

$$PD = Poss\left(A \succeq B\right) = \sup_{x,y:x\geq y} \min\left\{\mu_A\left(x\right), \mu_B(y)\right\}. \tag{15}$$

PD is the measure for a possibility that the set A is not dominated by the set B. In the considered problem of testing hypotheses we have (see Hryniewicz [11])

$$PD = Poss\left(\widetilde{H} \succeq \widetilde{K}\right) = \max\left\{\mu_{p_H}\left(0\right), \mu_{p_K}\left(1\right)\right\}, \tag{16}$$

and PD represents a *possibility* that choosing H over K is not a worse solution.

Second measure is called the *Possibility of Strict Dominance*, and for two fuzzy sets A and B is defined as

$$PSD = Poss\left(A \succ B\right) = \sup_{x} \inf_{y:y\geq x} \min\left\{\mu_A\left(x\right), 1 - \mu_B\left(y\right)\right\}. \tag{17}$$

PSD is the measure for a possibility that the set A strictly dominates the set B. In the considered problem of testing hypotheses we have

$$PSD = Poss\left(\widetilde{H} \succ \widetilde{K}\right) = \min\left\{\mu_{p_H}\left(0\right), 1 - \mu_{p_K}\left(0\right)\right\}, \tag{18}$$

and PSD represents a *possibility* that choosing H over K is a correct decision.

Third measure, called the *Necessity of Strict Dominance*, has been already introduced in the second section of this paper, and for two fuzzy sets A and B is defined by (5). The NSD index is related to the PD index in the following way

$$NSD = Ness\left(A \succ B\right) = 1 - Poss\left(B \succeq A\right), \tag{19}$$

and represents a *necessity* that the set A strictly dominates the set B. In the considered problem of testing hypotheses we have

$$NSD = Ness\left(\widetilde{H} \succ \widetilde{K}\right) = 1 - \max\left\{\mu_{p_H}\left(1\right), \mu_{p_K}\left(0\right)\right\}, \tag{20}$$

and NSD represents a strict *necessity* of choosing H over K.

Close examinations of the proposed measures reveals that for all possibility distributions defined by (13) and (14) holds the following relation

$$PD \geq PSD \geq NSD. \tag{21}$$

It means that according to the practical situation we can choose the appropriate measure of the correctness of our decision. If the choice between H and K leads to serious consequences we should choose the NSD measure. In such a case $p_H > 0,5$ is required to have $NSD > 0$. When these consequences are not so serious we may choose the PSD measure. In that case $PSD > 0$ when $p_K < 0,5$, i.e. when there is no strong evidence that the alternative hypothesis is true. Finally, the PD measure gives us the information of the possibility that choosing H over K is not a wrong decision.

Another interpretation of the possibility and necessity indices can be found in the framework of the recently rapidly developing theory of preference relations. Let $\mu(x,y)$ be a measure of the preference of x over y. The preference relation is complete (see [4] for further references) when

$$\mu(x,y) + \mu(y,x) \geq 1 \; \forall x, y \tag{22}$$

For a complete set of preference relations we may define the measure of the indifference between the alternatives x and y

$$\mu_I(x,y) = \mu(x,y) + \mu(y,x) - 1, \tag{23}$$

the measure that the alternative x is better than y

$$\mu_B(x,y) = \mu(x,y) - \mu_I(x,y), \tag{24}$$

and the measure that the alternative x is worse than y

$$\mu_W(x,y) = \mu(y,x) - \mu_I(x,y). \tag{25}$$

For the possibility and necessity measures defined above it is easy to show that

$$\mu_I(x,y) = PD - NSD \tag{26}$$

$$\mu_B(x,y) = NSD \tag{27}$$

$$\mu_W(x,y) = 1 - PD \tag{28}$$

Thus, the knowledge of the indices PD and NSD is sufficient for the complete description of preference relations for two alternatives: to choose the null hypothesis H or to choose the alternative hypothesis K.

7 Conclusions

In this paper we propose another way of looking at the results of statistical tests. The proposed method allows a practitioner to evaluate the test results using intuitive concepts of possibility, necessity or indifference. The proposed method allows to use a language that is closer to a common language used by practitioners than the specialised language of statistics. This might be very valuable, especially for computerised decision support systems that are oriented on users who are not well trained in statistics.

The main result of the paper is the new possibilistic interpretation of the well known concept of the observed test size p (also known as *p-value* or *significance*). This new interpretation has allowed to generalise the concept of the observed test size p for the case of imprecisely defined statistical hypotheses and vague statistical data.

For the evaluation of a single test we use the Necessity of Strict Dominance (NSD) index. This index generalises the concept of the observed test size p. In the case of fuzzy statistical hypotheses and fuzzy statistical data it is possible, however, to consider other possibilistic measures for the evaluation of the results of statistical tests. The interpretation of these indices, and their practical application, could be the subject of future research.

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, vol.75 (1995), 365–372.

[2] Arnold B. F., *An Approach to Fuzzy Hypothesis Testing*, Metrika, vol. 44 (1996), 119–126.

[3] Casals R., Gil M.A., Gil P., *The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information*, Europ. Journ. of Oper. Res., vol.27 (1986), 371–382.

[4] Cutell V., Montero J., *An Extension of the Axioms of Utility Theory Based on Fuzzy Rationality Measures*, In: Preference and Decisions under Incomplete

Knowledge, J.Fodor, B.De Baets, P.Perny (Eds.), Physica-Verlag, Heidelberg (1999), 33–50.

[5] Dubois D., Prade H., *Ranking fuzzy numbers in the setting of possibility theory*, Information Sciences, vol.30 (1983), 184–244.

[6] Dubois D., Prade H., *Fuzzy sets and statistical data*, Europ. Journ. of Oper. Res., vol.25 (1986), 345–356.

[7] Dubois D., Foulloy L., Mauris G., Prade H., *Probability-possibility transformations, triangular fuzzy-sets and probabilistic inequalities*, Proc. of the Ninth International Conference IPMU, Annecy (2002), 1077–1083.

[8] Grzegorzewski P., *Testing statistical hypotheses with vague data*, Fuzzy Sets and Systems, vol.112 (2000), 501–510.

[9] Grzegorzewski P., Hryniewicz O. *Testing hypotheses in fuzzy environment*, Mathware and Soft Computing, vol.4 (1997), 203–217.

[10] Grzegorzewski P., Hryniewicz O. *Soft Methods in Hypotheses Testing*, In: Soft Computing for Risk Evaluation and Management. Applications in Technology, Environment and Finance, D. Ruan, J.Kacprzyk, M.Fedrizzi (Eds.), Physica Verlag, Heidelberg (2001), 55–72.

[11] Hryniewicz O. *Possibilistic interpretation of the results of statistical tests*, Proc. of the Eight International Conference IPMU, Madrid, vol.I (2000), 215–219.

[12] Kruse R., *The strong law of large numbers for fuzzy random variables*, Information Sciences, vol.28 (1982), 233–241.

[13] Kruse R., Meyer K.D. *Statistics with Vague Data*, Riedel, Dodrecht, 1987.

[14] Kwakernaak H., *Fuzzy random variables, part I: definitions and theorems*, Information Sciences, vol.15, 1–15; *Part II: algorithms and examples for the discrete case*, Information Sciences, vol.17, 253–278.

[15] Lehmann E.L., *Testing Statistical Hypotheses*, 2nd ed., J.Wiley, New York (1986).

[16] Montenegro M., Casals M.R., Lubiano M.A., Gil M.A., *Two-sample hypothesis tests of means of a fuzzy random variable*, Information Sciences, vol.133 (2001), 89–100.

[17] Römer Ch., Kandel A., *Statistical tests for fuzzy data*, Fuzzy Sets and Systems, vol.72 (1995), 1–26.

[18] Saade J., *Extension of fuzzy hypothesis testing with hybrid data*, Fuzzy Sets and Systems, vol.63 (1994), 57–71.

[19] Saade J., Schwarzlander H., *Fuzzy hypothesis testing with hybrid data*, Fuzzy Sets and Systems, vol.35 (1990), 197–212.

[20] Son J.Ch., Song I., Kim H.Y., *A fuzzy decision problem based on the generalized Neyman-Pearson criterion*, Fuzzy Sets and Systems, vol.47 (1992), 65–75.

[21] Taheri S.M., Behboodian J., *Neyman-Pearson Lemma for Fuzzy Hypotheses Testing*, Metrika, vol.49 (1999), 3–17.

[22] Watanabe N., Imaizumi T. *A fuzzy statistical test of fuzzy hypotheses*, Fuzzy Sets and Systems, vol.53 (1993), 167–178.

[23] Zadeh L.A., *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets and Systems, vol.1 (1978), 3–28.