International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Statistical properties of the fuzzy p-value ☆

Olgierd Hryniewicz

Systems Research Institute, Polish Academy of Sciences, Newelska 6, Warszawa, Poland

ARTICLE INFO

Article history:
Received 29 December 2016
Received in revised form 6 November 2017
Accepted 11 December 2017
Available online xxxx

Keywords:
Statistical tests
Fuzzy data
Fuzzy hypotheses
Fuzzy p-value

ABSTRACT

We have considered different approaches for the calculation of the p-value for fuzzy statistical tests. We have considered several approaches known from literature, and a new one, based on Lehmann's concept of the interval statistical hypothesis. For the particular problem of testing hypotheses about the mean in the normal distribution with known standard deviation, and a certain type of fuzziness (both in data and tested hypotheses), we have found approximate probability distributions of the respective defuzzified p-values. These distributions let us evaluate the compatibility of the observed data with the assumed hypothetical model.

1. Introduction

The concept of p-value is probably the most frequently used and misused concept of statistics. It was formally introduced by R.A. Fisher in the 1920's, but practically it was used earlier, e.g., in works of Karl Pearson. Its usage represents an inductive approach to statistical data analysis. In this approach, the whole statistical analysis is based on one particular sample of observations, and no other concepts, like, e.g., potential repetition of experiments, are needed. One can even say that in this approach the concept of the classical frequency-based probability is not needed. What Fisher really does need, is the concept of an infinitely large population, and his inductive statistical inference is focused on the rejection of a null hypothesis about this population. In contrast to this Fisherian approach, the approach proposed by Neyman and Pearson is deductive, and strongly based on the frequentist concept of probability. They claim, however, that their theory represents "inductive behavior" in establishing a rule for making decisions about two hypotheses: null and alternative. Therefore, even for the same statistical data decisions made using these two approaches may be (and usually are) different, and their interpretations must also be different.

Many practitioners, trying to interpret results of their experiments, mix these two approaches, and arrive at completely false conclusions. The situation is even more complicated when we take into account the third paradigm of data analysis – the Bayesian one. In this approach – in its "objective" version proposed by Jeffreys – an uninformative prior distribution defined on the parametric set is introduced, and then a posterior probability distribution, conditioned on the observed sample, is calculated. The decision in favor of the null hypothesis is taken if the quotient of the posterior probabilities of the null and the alternative hypotheses is larger than a certain critical value. This third approach is important in this sense, as it provides a formal platform where both deductive and inductive approaches can meet, because only in the Bayesian setting we can say about the probability that the tested hypothesis is true. This problem will be addressed in the next section of

☆ This paper is part of the Virtual special issue on Soft methods in probability and statistics, Edited by Barbara Vantaggi, Maria Brigida Ferraro, Paolo Giordani.

E-mail address: hryniewi@ibspan.waw.pl.

https://doi.org/10.1016/j.ijar.2017.12.003
0888-613X/© 2017 Published by Elsevier Inc.

this paper. More information about distinctive features of different approaches to statistical testing using the concept of p-value can be found in numerous papers, such as, e.g., [2] or [6].

The controversies between different approaches to the interpretation of statistical tests are even amplified when we consider the problem of statistical testing in a fuzzy statistical environment. By a fuzzy statistical environment we understand situation when both statistical data and/or statistical hypotheses can be imprecisely perceived or defined. Statistical tests used in this environment are usually called fuzzy statistical tests. The problems with the understanding of fuzzy statistical tests begin with the used interpretation of the concept of a fuzzy random variable. Depending on "epistemic" or "ontic" interpretation of fuzzy random data (see [7] for more information) the interpretation of the results of fuzzy statistical tests may be quite different.

The concept of p-value is absolutely crucial in practical statistical decision making. The reported p-value is often the only information provided by statistical packages and statistical functions of popular spreadsheets, like MS Excel. Even if a statistical package provides more information (e.g., in form of confidence intervals), the majority of practitioners use only the concept of p-value in their decision making processes. In many areas of applied science, such as medicine, biology or environmental sciences, the p-values must be reported while presenting the results of experiments. Therefore, the analysis of consequences of the usage of p-values in the fuzzy statistical environment seems to be of important practical value.

The paper is a significant extension of the conference paper by Hryniewicz [25]. More models of fuzzy statistical tests are considered, and better approximations for probability distributions of considered statistics have been proposed. The rest of the paper is organized as follows. In the second section, we present the concept of p-value, as it is used and interpreted in classical statistics. We begin with some historical remarks, and then we present mathematical description of the concept, and finally present its often disputable interpretation. In the third section, we present different approaches to the usage of the concept of p-value in fuzzy statistical tests. We discuss the application of differently defined p-values that can be used in a fuzzy statistical environment. This discussion is illustrated with some examples of testing fuzzy statistical hypotheses. In Section 4, using extensive computer simulation experiments, we analyze probability distributions of different defuzzified versions of fuzzy p-values in the case of statistical tests about the expected value of the normal distribution. The paper is concluded in its last section.

2. The concept of p-value – crisp case

The concept of p-value has been defined in many ways. Below, following [33], we present a definition that can be used in further considerations. Let $\mathbf{X}$ be random data described by a continuous density function $f(\mathbf{x})$. Let us also assume that in our decision model this density is completely specified, and forms our simple hypothetical model H_0 about a parameter θ. Compatibility of this hypothetical model with observed data $\mathbf{x}$ is evaluated using a certain statistic $T(\mathbf{X})$ whose large values indicate less compatibility. The p value is then defined as

$$p(x) = P_\theta(T(\mathbf{X}) \geq T(\mathbf{x})). \tag{1}$$

In a more general setting, the distribution of $\mathbf{X}$ may be absolutely continuous wrt the Lebesgue measure, and the tested null hypothesis may be not simple, and defined as $\theta \in \Theta_0$. In this case a more general definition of the p-value is given by

$$p(x) = \sup_{\theta \in \Theta_0} P_\theta(T(X) > T(x)). \tag{2}$$

Let us reformulate the definition (1) making it more understandable, but in some cases more difficult for computation. Let M be our hypothetical model (parameter of a distribution, cumulative probability function, etc.), and $M_x(X)$ be a sample statistic that describes M. Let $d(X) = d(M_x(X) - M)$ be a non-negative function that measures appropriately defined "distance" between $M_x(X)$ and M. Then, the definition (1) may be reformulated as

$$p = P(T'(d(X)) \geq T'(d(x))). \tag{3}$$

This representation is especially useful for testing hypotheses about a parameter of location. For example, in testing hypotheses about the expected value of a normally distributed random variable with known value of σ, such that $\sigma\sqrt(n) = 1$, we have the following simple formulae for the calculation of p-values [32]:

$$p_l = \Phi(\mu_0 - \bar{x}), \tag{4}$$

for testing the one-sided null hypothesis $H_0 : \mu \leq \mu_0$ against the alternative $H_A : \mu > \mu_0$,

$$p_u = \Phi(\bar{x} - \mu_0), \tag{5}$$

for testing the one-sided null hypothesis $H_0 : \mu \geq \mu_0$ against the alternative $H_A : \mu < \mu_0$, and

$$p_u = 2 * \Phi(-|\bar{x} - \mu_0|), \tag{6}$$