Approximated Fuzzy p-values by Bootstrapped Fuzzy Distributions and Fuzzy Hypotheses Testing

Julien Rosset[1] [a], Laurent Donzé[1] [b]

[1]*ASAM Group, University of Fribourg, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland*
{*Julien.Rosset, Laurent.Donzé*}*@unifr.ch*

Keywords: Fuzzy confidence interval; Fuzzy hypothesis testing procedure; Fuzzy p-value; Fuzzy logic; Fuzzy methods in data analysis.

Abstract: Although we could dispute the use of p-values, it is a standard tool used by many to know if one has to reject or not a null hypothesis. With the emergence of fuzzy data, fuzzy hypothesis testing procedures appeared. Along these testing procedures, various methods to compute crisp or fuzzy p-values arising from fuzzy data and hypotheses were investigated. However, we noticed that, despite calculating a fuzzy test statistic, none of these approaches assume a fuzzy distribution for it. Thus, to remedy to this problem, we tackle the problem of finding fuzzy p-values in the context of both fuzzy data and hypotheses while assuming a fuzzy distribution for the test statistic. Finding fuzzy p-values alone is not useful if one does not know how to use them to take a decision. This is why we also provide a way to interpret fuzzy p-values and present a decision rule as to when one should reject or not the fuzzy null hypothesis. Additionally, we aim at comparing this decision rule to fuzzy statistical testing procedures. To meet this objective, we offer an empirical application that compares the obtained fuzzy p-values to the results given by a fuzzy hypothesis testing procedure. This gives us the possibility to thoroughly discuss the differences between the two approaches. This paper thus proposes a method to calculate fuzzy p-values assuming a fuzzy distribution for the test statistic and explain how to interpret and take a decision with them. It also compares the decision taken with fuzzy p-value to the one taken with a fuzzy hypothesis testing procedure.

1 INTRODUCTION AND NOTATION

While it is debatable whether or not the p-value is a good tool for testing statistical hypotheses, it is widely used to reject or not the null hypothesis. It has been decades that researchers are extending classical hypotheses testing procedures. There are three possible extensions: (1) using fuzzy data ; (2) using fuzzy hypotheses ; (3) assuming a fuzzy distribution for the test statistic. Several authors already extended the first two. To cite a few, Viertl (Viertl, 2018) investigated the problem of finding fuzzy p-values in the context of fuzzy hypotheses and crisp data. (Parchami, 2020) and (Berkachy and Donzé, 2022) developed a method to find fuzzy p-values with both fuzzy data and hypotheses. While many methods to compute p-values and fuzzy p-values already exist, the assumption of a crisp distribution for the sample out of which the test statistic is computed is widely used to simplify the calculation of fuzzy p-values. Thus, regarding (3), to our knowledge, no one has yet considered it. This is why, based upon the previously developed methods, this paper aims at meeting the following objectives. (a) Finding fuzzy p-values for both fuzzy data and hypotheses assuming a fuzzy distribution for the test statistic, that is, a procedure extending all three points mentioned above. (b) Fuzzy p-values obtained by this procedure must also reduce to classical p-values when the data and hypotheses are crisp. (c) Providing a decision rule and an interpretation of these fuzzy p-values. Additionally, through a numerical application, a comparison between decisions taken with fuzzy p-values and a fuzzy hypothesis testing procedure is presented.

On other occasions, we had the opportunity to present fuzzy test statistics. To compare the results of inferences from fuzzy p-values and those from our fuzzy test statistics, we will briefly recall these approaches in Section 4. Finally, we aim to investigate the limits of the inference procedure based ex-

[a] https://orcid.org/0000-0002-7883-1512
[b] https://orcid.org/0000-0003-3522-4672

clusively on fuzzy p-values. To achieve this, we provide two case studies to compare the results based on fuzzy p-values to those obtained via a fuzzy hypothesis testing procedure and discuss them thoroughly.

We recall in Section 2 the main recent developments on the subject of fuzzy inferences. Then, we explain in Section 3 our method to compute p-values and propose a decision rule to interpret them. A small overview of the developed fuzzy tests by Rosset and Donzé (Rosset and Donzé, 2024a) is given in Section 4. Then, in Section 5, two applications with data permit us to illustrate our methods. Finally, we conclude in Section 6.

We will use the following notation and conventions below. First, let us define by $\tilde{x}$ a fuzzy number and by $\mu_{\tilde{x}}(\cdot)$ its membership function. We consider also the α-cuts of $\tilde{x}$ denoted by $\tilde{x}^{\alpha}$ or by its equivalent in interval form by $(x^{L,\alpha}, x^{R,\alpha})$. In practice, triangular fuzzy numbers are often used. We denote them by a triplet $\tilde{x} = (x^L, x, x^R)$ with $x^L \leq x \leq x^R \in \mathbb{R}$. Furthermore, assuming a random sample $X_1, \dots, X_n$ with its realisations $x_1, \dots, x_n$, we note by $\tilde{x}_1, \dots, \tilde{x}_n$ the fuzzy equivalent of these quantities and by $\tilde{x}_i^{\alpha}$, $i = 1, \dots, n$ their α-cuts.

2 LITERATURE REVIEW

Oftentimes, with the study of inference comes the study of p-values. It is therefore natural to analyse the p-values arising from these fuzzy testing procedures. Let us recall the three key elements that define a classical procedure to find p-values: the use of crisp data; the use of crisp hypotheses; due to the crisp nature of the data, the use of a crisp distribution describing the sample out of which the test statistic is computed. Thus, such procedures can be extended to fuzzy procedures in different ways: (1) using fuzzy hypotheses with crisp data; (2) using crisp hypotheses with fuzzy data; (3) using both fuzzy hypotheses and data; (4) in the presence of fuzzy data, assuming a fuzzy distribution of the sample. (1)-(3) have already been thoroughly studied by researchers, but (4) has, to our knowledge, not yet been considered. Regarding (1)-(3), (Viertl, 2018) investigated fuzzy p-values coming from fuzzy data with crisp hypotheses. To do so, he used a fuzzy test statistic and assumed a crisp distribution for the sample. This setting allows him to compute a p-value corresponding to each of the left and right alpha-cuts of the test statistic. This collection of p-values then corresponds to the alpha-cuts of a fuzzy p-value. As opposed to Viertl, (Parchami et al., 2018) analysed the case where hypotheses are fuzzy and data are crisp. They introduced the notion of fuzzy p-value by applying Zadeh's extension principle. Later, (Parchami, 2020) investigated the case where both hypotheses and data are fuzzy and studied the fuzzy p-values arising from Zadeh's extension principle under the assumption that the test statistic has a crisp distribution despite the fuzziness of the data. Similarly, (Berkachy and Donzé, 2017) and (Berkachy, 2020) presented a method to find fuzzy p-values under the condition of fuzzy data and hypotheses assuming a crisp distribution of the sample. We can also mention (Hryniewicz, 2018) who developed a procedure to find fuzzy p-values using fuzzy confidence intervals in the sense of (Kruse and Meyer, 1987). His method is based on finding intersection points between the fuzzy null hypothesis and fuzzy confidence interval of the test statistic. Despite being a very interesting method, the procedure has two problems. First, it doesn't deal with a fuzzy distribution of the test statistic. Secondly, when the fuzzy confidence interval becomes crisp, the intersection points (p-values) take the value 0 or 1, thus not generalising the classical approach to find p-values.

As mentioned above, the study of fuzzy p-values involves fuzzy hypothesis testing procedures, which are connected to the notion of fuzzy confidence intervals. Hence, let us briefly review some results in this context. (Kruse and Meyer, 1987) gave a theoretical definition of a fuzzy confidence interval. Many researchers have utilised and refined their definition since its introduction in 1987. Indeed, (Viertl and Yeganeh, 2016) introduced the notion of fuzzy confidence regions. (Kahraman et al., 2016), (Kahraman et al., 2019) studied interval-valued intuitionistic fuzzy sets (IVIFSs) and hesitant fuzzy sets (HFS) and, based on them, developed two methods to construct fuzzy confidence intervals. (Wu, 2009) solved optimisation problems involving a fuzzy Gaussian distribution to build fuzzy confidence intervals.(Chachi and Taheri, 2011) proposed fuzzy confidence intervals for the mean of a fuzzy normal distribution. These approaches involve a pre-defined distribution around a specific parameter value, which, in practical cases, is not always possible to know. To bypass this difficulty, (Berkachy and Donzé, 2022) developed a general procedure to construct fuzzy confidence intervals using the likelihood ratio method and a bootstrap procedure extended to the fuzzy environment to estimate the distribution of this likelihood ratio.

Regarding bootstrap techniques, we can cite the work of Bradley Efron popularised bootstrap methods among statisticians. Then, many fuzzy bootstrap approaches were introduced. Among the recent ones, we can point out (Berkachy and Donzé, 2020), who gave two algorithms to generate bootstrapped samples with the notions of location and dispersion and (Grzegorzewski and Romaniuk, 2021), who provided an epistemic approach to fuzzy bootstrap techniques for fuzzy data.

Finally, let us mention two concepts that come in handy in constructing fuzzy confidence intervals. The first one is the fuzzy quantile function proposed by (Shvedov, 2016), a fuzzy extension of the classical quantile function. The second one is the construction of fuzzy distributions where (Arefi et al., 2012) describe how to find empirical fuzzy distributions.

3 FUZZY P-VALUES AND INFERENCES

Statistical inferences are based on samples. Let us denote by $X_1,\ldots,X_n$ a random sample drawn from the distribution of interest. Assuming the data are fuzzy, we write as $\tilde{X}_1,\ldots,\tilde{X}_n$, the fuzzy equivalent of the random sample, and as $\boldsymbol{X} = (\tilde{X}_1,\ldots,\tilde{X}_n)$ the fuzzy vector with its membership function $\mu_{\tilde{X}}$, such that $\mu_{\tilde{X}} : \mathbb{R}^n \to [0,1]^n$. We note by $\tilde{\bar{X}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{X}$ the fuzzy random mean and its realisation by $\tilde{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i$.

It's common practice to express a fuzzy hypothesis using linguistic terms. For example, we could say, "*The value of the parameter* θ *is more or less equal to* θ_0", or rewritten in terms of differences, i.e. "*The difference between the value of the parameter* θ *and* θ_0 *is more or less equal to 0*". Thus, we could write, for instance the following two-sided fuzzy hypotheses:

$$\tilde{H}_0 \ : \ \theta = \tilde{0} \ \text{ against } \ \tilde{H}_1 \ : \ \theta \neq \tilde{0}. \tag{1}$$

Of course, in the same way, one-sided fuzzy hypotheses could be written as:

$$\tilde{H}_0 \ : \ \theta \leq \tilde{0} \ \text{ against } \ \tilde{H}_1 \ : \ \theta > \tilde{0}, \tag{2}$$

or

$$\tilde{H}_0 \ : \ \theta \geq \tilde{0} \ \text{ against } \ \tilde{H}_1 \ : \ \theta < \tilde{0}. \tag{3}$$

Remark that the null fuzzy hypothesis $\tilde{H}_0$ or the alternative one $\tilde{H}_1$ can be model by triangular numbers, i.e. (H_0^L, H_0, H_0^R). On the other hand, it is common practice to implement the inference process through a proper test statistic. Let $\tilde{T} = (T^L, T, T^R)$ be a fuzzy statistic to test $\tilde{H}_0$. For example, if we want to test the mean of the distribution, we could use:

$$\tilde{T} := (\tilde{\bar{X}} - \theta_0) \oslash \tilde{\sigma}_{\tilde{\bar{x}}}, \tag{4}$$

where $\oslash$ is the fuzzy division operator and $\tilde{\sigma}_{\tilde{\bar{x}}}$ is the standard-error of $\tilde{\bar{X}}$. We note by $\tilde{t} = (t^L t, t^R)$ the fuzzy realisation of $\tilde{T}$.

In a fuzzy approach, the distribution of the statistic is difficult to know, even more so if one doesn't want to assume strong conditions about the form of the distribution. As we need this distribution, we propose using a bootstrapped one instead. One of the many advantages of using this approach is that it is relatively easy to generate a fuzzy bootstrapped distribution for $\tilde{T}$. Let $\tilde{t}_1^\star,\ldots,\tilde{t}_b^\star,\ldots\tilde{t}_B^\star$ be the bootstrapped sample yielding the test statistic's distribution and let $\tilde{\bar{t}}^* = (\sum_1^B \tilde{t}_b^\star)/B = (\bar{t}^{*L}, \bar{t}^*, \bar{t}^{*R})$ be the associated fuzzy empirical mean.

We consider, on one hand, the fuzzy observed statistic $\tilde{t} = (t^L, t, t^R)$ and its absolute value:

$$\begin{aligned}
|\tilde{t}| =& (t_{abs}^L, t_{abs}, t_{abs}^R) \\
=& \big(\min(|t^L|, |t|, |t^R|), \\
& \quad |t|, \\
& \quad \max(|t^L|, |t|, |t^R|) \big),
\end{aligned}$$

and, on the other hand, the centred bootstrapped distribution given by $\tilde{t}_1^\star - \bar{t}^*,\ldots,\tilde{t}_b^\star - \bar{t}^*,\ldots\tilde{t}_B^\star - \bar{t}^*$. We note this distribution by $\tilde{F}^*$. Then, we can compute the following empirical approximation for the fuzzy p-value $\tilde{p} = (p^L, p, p^R)$ of the two-sided hypotheses test (1),

$$\begin{aligned}
p^L &= 2 \cdot \frac{\#(F^{\star,L} > t_{abs}^R) + 1}{B+1} \\
p &= 2 \cdot \frac{\#(F^\star > t_{abs}) + 1}{B+1} \\
p^R &= \min(2 \cdot \frac{\#(F^{\star,R} > t_{abs}^L) + 1}{B+1}, 1),
\end{aligned}$$

where # gives the number of cases for which the condition in parentheses is true. For the one-sided hypotheses test (2), the fuzzy p-value $\tilde{p} = (p^L, p, p^R)$ is calculated as

$$\begin{aligned}
p^L &= \frac{\#(F^{\star,L} > t^R) + 1}{B+1} \\
p &= \frac{\#(F^\star > t) + 1}{B+1} \\
p^R &= \min(\frac{\#(F^{\star,R} > t^L) + 1}{B+1}, 1),
\end{aligned}$$

and for the second one-sided hypotheses test (3), we have

$$p^L = \frac{\#(F^{\star,R} < t^L) + 1}{B+1}$$
$$p = \frac{\#(F^{\star} < t) + 1}{B+1}$$
$$p^R = \min(\frac{\#(F^{\star,L} < t^R) + 1}{B+1}, 1).$$

We must underline that we centred the fuzzy bootstrapped distribution (left, centre and right parts) around $\tilde{t}^*$. Another possibility would've been to centre it around $\bar{\tilde{t}}^*$ by writing it as $\tilde{t}_1^{\star} \ominus \bar{\tilde{t}}^*, \dots, \tilde{t}_b^{\star} \ominus \bar{\tilde{t}}^*, \dots \tilde{t}_B^{\star} \ominus \bar{\tilde{t}}^*$ where $\ominus$ is the fuzzy substraction operator. Nevertheless, we did not choose it for empirical reasons. The algorithm 1 summarises the computation steps to find the bootstrapped fuzzy p-values.

Data: $\tilde{x}_1, \dots, \tilde{x}_n$ a fuzzy realisation of the random sample.
Result: Fuzzy p-values.
begin
  Compute: ($\tilde{t}$, $|\,\tilde{t}\,|$) ;
  for $b = 1$ *to* B **do**
    Draw sample $\tilde{s}_b = \{\tilde{x}_1^*, \dots, \tilde{x}_n^*\}$;
    Compute: ($\tilde{t}_b^*$) ;
  end
  Compute: ($\bar{\tilde{t}}^*$, $\tilde{F}^*$, $\tilde{p}$) ;
  return $\tilde{p}$;
end

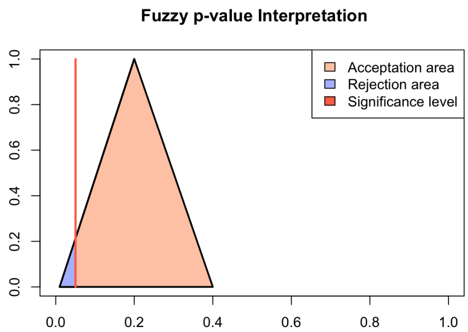
Algorithm 1: Fuzzy p-values by a Bootstrapped Fuzzy Distribution.

Fuzzy p-values $\tilde{p} = (p^L, p, p^R)$ are not always easy to interpret, unlike their crisp counterparts. Though a proper defuzzification method can help the interpretation, we propose constructing a specific measure whose meaning is straightforward. Let $\delta \in (0,1)$ be a chosen significance level. We expect $\delta \in [p^L, p^R]$, but this could not be true. Let us suppose that this is the case. Furthermore, we assume first that $\delta \leq p$. In the space $[0,1]^2$, consider the point (δ, y_0) given by the intersection between the line $x = \delta$ and the triangular fuzzy number (p^L, p, p^R) and compute the area A of the new resulting triangle given by the three points $((p^L, 0), (\delta, y_0), (\delta, 0))$. If $\delta > p$, then A is the area of the polygon given by the four points $((p^L, 0), (p, 1), (\delta, y_0), (\delta, 0))$. On the other hand, let P be the area of the fuzzy p-value. Then, we can propose the following crisp measure of the p-value:

$$r^{rej} = A/P. \tag{5}$$

We set $r^{rej} = 0$ if $\delta < p^L$ and $r^{rej} = 1$ if $\delta > p^R$. The real number r^{rej} is between 0 and 1 and tells us how much the fuzzy p-value tends to reject the null hypothesis. On the opposite, $r^{acc} = 1 - r^{rej} \in [0,1]$ gives us the tendency of not rejecting the null hypothesis. If $r^{rej} > 0.5$, we will tend to reject the null hypothesis, while if $r^{rej} < 0.5$, we will tend not to reject it. No decision can be taken for $r^{rej} = 0.5$. r^{rej} and r^{acc} can be seen as weights to respectively reject or not to reject the null hypothesis. Figure 1 depicts the tendency to reject the null hypothesis in blue. This area equals to $r^{rej} = 0.22$. The orange region has an area of $r^{acc} = 1 - r^{rej} = 0.78$ and gives the tendency not to reject the null hypothesis.

Figure 1: Example of a fuzzy p-value. Here $\tilde{p} = (p^L, p, p^R) = (0.01, 0.2, 0.4)$. The tendency to reject the null hypothesis (in blue) is $r = 0.22$, and the tendency to not reject the null hypothesis (in orange) is $1 - r = 0.78$



4 Fuzzy tests and confidence intervals

Rosset and Donzé (Rosset and Donzé, 2024a) proposed the following inference procedure. They built two real-valued functions of the fuzzy random sample $\tilde{X}_1, \dots, \tilde{X}_n$ defined as

$$\tilde{\phi}^{\mathrm{rej}} : (\mathbb{F}_c^*(\mathbb{R}))^n \to [0,1], \qquad \tilde{\phi}^{\mathrm{acc}} : (\mathbb{F}_c^*(\mathbb{R}))^n \to [0,1].$$

It can be proven that $\tilde{\phi}^{\mathrm{rej}}$ and $\tilde{\phi}^{\mathrm{acc}}$ are two fuzzy statistical tests on $\mathbb{F}_c^*(\mathbb{R})$, the class of the non-empty compact, convex and normal fuzzy sets on $\mathbb{R}$. Functions of the random sample X, their values are given by the intersection of the fuzzy null hypothesis $\tilde{H}_0$ and a fuzzy confidence interval $\tilde{\Pi}$ for the parameter to be tested θ. For technical details, the reader may consult Rosset and Donzé (Rosset and Donzé, 2024a). The interpretation of these statistical tests is given in the following definition:

Definition 1 (Decision rules)**.**

Let be

$$L^{rej} = \int_{[0,1]} \tilde{\phi}^{rej,\alpha} d\alpha \qquad \textit{and} \qquad L^{acc} = \int_{[0,1]} \tilde{\phi}^{acc,\alpha} d\alpha. \tag{6}$$

L^{rej} and L^{acc} are called, respectively, the levels of rejection and acceptance (not rejection) of the (fuzzy) null hypothesis $\tilde{H}_0$.

1. *If $L^{rej} = 1$, $\tilde{H}_0$ is rejected. In this case $L^{acc} = 0$;*
2. *If $L^{rej} = 0$, $\tilde{H}_0$ is not rejected. In this case $L^{acc} = 1$;*
3. *If $0 < L^{rej} < 1$, the test shows how strong $\tilde{H}_0$ is rejected. The higher L^{rej} is, the stronger the test rejects $\tilde{H}_0$;*
4. *If $0 < L^{acc} < 1$, the test shows how strong $\tilde{H}_0$ is not rejected. The higher L^{acc} is, the stronger the test doesn't reject $\tilde{H}_0$;*
5. *If $L^{rej} > L^{acc}$, $\tilde{H}_0$ tends to be rejected;*
6. *If $L^{rej} < L^{acc}$, $\tilde{H}_0$ tends to be not rejected;*
7. *If $L^{rej} = L^{acc}$, no decision can be taken.*

The fuzzy confidence interval $\tilde{\Pi}$ for the parameter θ of the distribution, defined at a given confidence level $1-\delta$, plays a crucial role in the construction of the two fuzzy tests $\tilde{\phi}^{rej}$ and $\tilde{\phi}^{acc}$. Hence, its construction is a key point of the inference procedure. Several methods can be used to build such fuzzy confidence intervals. We will use the one proposed in Rosset and Donzé (Rosset and Donzé, 2024b), where we introduced a method based on bootstrap techniques. This method is attractive because no specific (fuzzy) distribution should be assumed.

5 APPLICATIONS

5.1 *Case 1*

Let us test the fuzzy mean of a fuzzy distribution at a significance level of 5%. Table 1 depicts the fuzzy observations taken from Berkachy and Donzé (Berkachy and Donzé, 2020) that will be used to compute the fuzzy p-values associated with different fuzzy null hypotheses as illustrated in Table 2. The fuzzy mean estimation is the triangular fuzzy number $\tilde{\bar{x}} = (1.8, 2.8, 3.8)$ and a fuzzy confidence interval at a confidence level of 95% is the fuzzy trapezoidal number $(0.933, 1.933, 3.6, 4.6)$. This fuzzy confidence interval will allow us to test the null hypotheses and compute their associated levels of acceptance and rejection as shown in Table 2. Lastly, we will compare the obtained fuzzy p-values to the results given by the fuzzy hypothesis testing procedure.

Let us first talk about the results given by the fuzzy hypothesis testing procedure at a 5% significant level in Table 2. One can easily see that in the extreme cases (tests 1, 2, 3, 13) when the fuzzy null hypothesis is outside the fuzzy confidence interval at 95% $(0.93, 1.93, 3.6, 4.6)$, the levels of acceptance and rejection are respectively 0 and 1. Moreover, when the fuzzy null hypothesis is entirely contained in the fuzzy confidence interval's core $(1.93, 3.6)$, e.g. tests 8, 9, 10, the levels of acceptance and rejection are respectively 1 and 0. These two extreme situations give the expected binary results one would have found using a crisp testing procedure. Let us now analyse the cases when the fuzzy null hypothesis overlaps with the fuzzy confidence interval's fuzzy regions, i.e. $(0.93, 1.93)$ and $(3.6, 4.6)$. With tests 4, 5, 6 and 7 of Table 2, when the null hypothesis is overlapping the interval $(0.93, 1.93)$, we observe that the values of L^{acc} and L^{rej} change depending on the shape of H_0. As opposed to the crisp case, we see that the values L^{acc} and L^{rej} are in between 0 and 1 and will tend to reject or not the null hypothesis. Tests 5 and 7 depicted on Figures 3 and 4 show how the shape of a crispier null hypothesis (Figure 4) or fuzzier null hypothesis (Figure 3) affects the values of L^{acc} and L^{rej}. Lastly, when the fuzzy confidence interval is fully overlapped (tests 11 and 12), the amount of area outside the fuzzy confidence interval will dictate the value of the rejection level, and conversely, the amount of area inside the fuzzy confidence interval will dictate the value of the acceptance level. This situation is illustrated in Figures 2 and 5.

Figure 2: Fuzzy test of the mean θ, $\tilde{H}_0 : \theta = (0, 3.8, 6)$ against $\tilde{H}_1 : \theta \neq (0, 3.8, 6)$. FH0 stands for "Fuzzy Null Hypothesis" and FCI stands for "Fuzzy Confidence Interval". The tendency to not reject the null hypothesis (acceptance area in purple on the left graph) is $L^{acc} = 0.68$. The tendency to reject the null hypothesis (rejection area in purple on the right graph) is $L^{rej} = 0.43$.

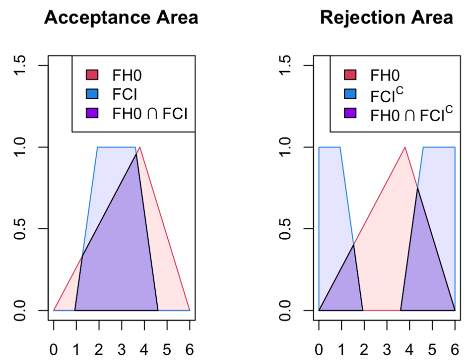


Figure 3: Fuzzy test of the mean θ, $\tilde{H}_0 : \theta = (1, 1.5, 2)$ against $\tilde{H}_1 : \theta \neq (1, 1.5, 2)$. FH0 stands for "Fuzzy Null Hypothesis" and FCI stands for "Fuzzy Confidence Interval". The tendency to not reject the null hypothesis (acceptance area in purple on the left graph) is $L^{acc} = 0.75$. The tendency to reject the null hypothesis (rejection area in purple on the right graph) is $L^{rej} = 0.58$.

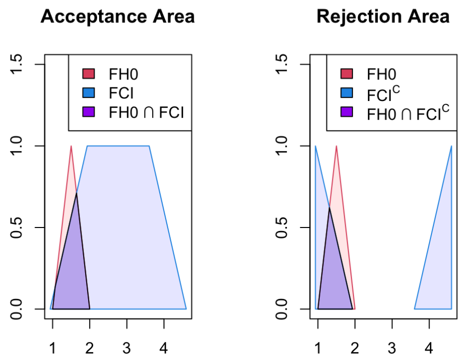
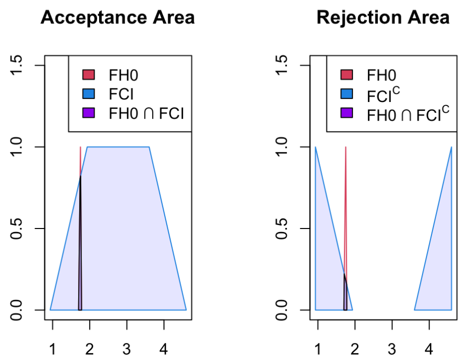
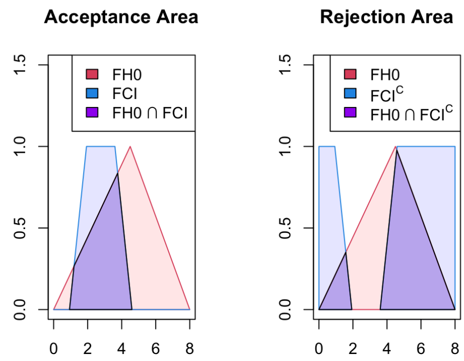


Figure 4: Fuzzy test of the mean θ, $\tilde{H}_0 : \theta = (1.7, 1.75, 1.78)$ against $\tilde{H}_1 : \theta \neq (1.7, 1.75, 1.78)$. FH0 stands for "Fuzzy Null Hypothesis" and FCI stands for "Fuzzy Confidence Interval". The tendency to not reject the null hypothesis (acceptance area in purple on the left graph) is $L^{acc} = 0.96$. The tendency to reject the null hypothesis (rejection area in purple on the right graph) is $L^{rej} = 0.34$.



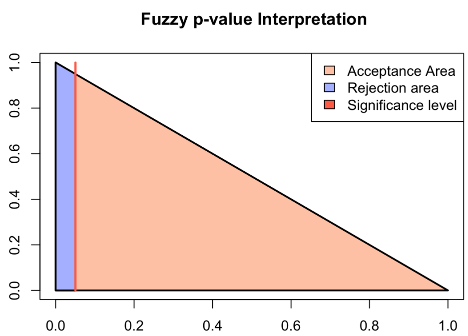
Let us now discuss the fuzzy p-values obtained by the method displayed in Section 3. In extreme cases, as in test 1 and 13, we find binary values. For test 1, we find the values, $r^{rej} = 0$, $r^{acc} = 1$ and for test 13, the values $r^{rej} = 1$, $r^{acc} = 0$. This shows that we tend to strongly reject (respectively not reject) the null hypothesis which confirms the decisions taken with L^{rej} and L^{acc}. However, one quickly notices that the fuzzy p-values tends to take extreme values for the left and right parts. Excepted tests 1, 12 and 13, the left and right parts of every fuzzy p-value are 0 and 1. This

Figure 5: Fuzzy test of the mean θ, $\tilde{H}_0 : \theta = (0, 4.5, 8)$ against $\tilde{H}_1 : \theta \neq (0, 4.5, 8)$. The tendency to not reject the null hypothesis (acceptance area in purple on the left graph) is $L^{acc} = 0.44$. The tendency to reject the null hypothesis (rejection area in purple on the right graph) is $L^{rej} = 0.63$.



makes the interpretation of these fuzzy p-values quite hard. One can, however, observe that when the null hypothesis is fully contained in the region delimited by $(1.93, 3.6)$, r^{rej} is very low, and r^{acc} is almost 1, which is in accordance with the results of the fuzzy testing procedure. However, in tests 2, 3 and 4, when the null hypothesis should be rejected, the fuzzy p-values give a r^{acc} of 0.9025 and a r^{rej} of 0.0975. While it seems surprising, it can be explained by looking at the meaning behind these two quantities. On Figure 6, One can see that although the left and centre

Figure 6: Extreme fuzzy p-value $\tilde{p} = (0, 0, 1)$.$r^{acc} = 0.9025$ and $r^{rej} = 0.0975$

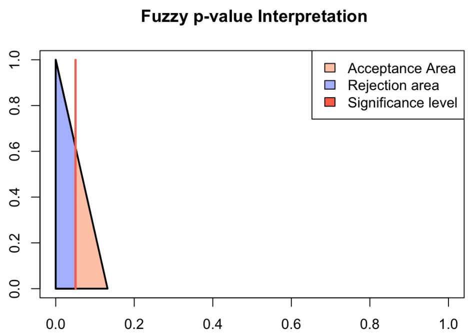


parts of the fuzzy p-value favour rejecting the null hypothesis (both have a value of 0), r^{acc} (orange area) is very large because the right part of the fuzzy p-value is 1. This causes the fuzzy p-value to have a great r^{acc} and low r^{rej}. This result shows the limits of the use of fuzzy p-values. However, let us point out the results for test 12 depicted in Figure 7. Observe that they are in accordance with the results found with the testing procedure. L^{acc} is 0.44 and r^{acc} is 0.39 which tends to reject the null hypothesis. This is confirmed by $L^{rej} = 0.63$ and $r^{rej} = 0.61$. This good result is a byproduct of having a fuzzy p-value with a small right part, which can be easily interpreted. Test 12 shows that when the fuzzy p-value is not too extreme, i.e. does not have a right part close to 1, it holds information about whether or not to reject the null hypothesis. To confirm this assumption, let us consider another set of crispier observations that will give us crispier p-values.

5.2 *Case 2*

Let $X \sim N(2.8,3)$, $U_1 \sim \text{Unif}(0.001,0.006)$, $U_2 \sim \text{Unif}(0.004,0.009)$. Let x_i, $u_{1,i}$, $u_{2,i}$ be realisations of the random variables X, U_1 and U_2 respectively. We can form fuzzy observations in the following way. Let $x_i^L = x_i - u_{1,i}$ and $x_i^R = x_i + u_{2,i}$. We generate N fuzzy triangular observations $\tilde{x}_i = (x_i^L, x_i, x_i^R)$, $i = 1,\ldots,N$. By generating $N = 500$ observations via the above procedure, we find an estimation for the fuzzy mean $\tilde{\bar{x}} = (2.693, 2.697, 2.704)$ and a fuzzy confidence interval at 95% confidence level $(2.44, 2.45, 2.94, 2.95)$. Table 5 displays fuzzy p-values obtained by a two-sided test for different null hypotheses for these 500 observations. Table 3 displays fuzzy p-values obtained by a one-sided test for the mean θ, $\tilde{H}_0 : \theta < \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \geq \tilde{\theta}_0$, where $\tilde{\theta}_0$ is the conjectured fuzzy triangular number used to characterise the null hypothesis $\tilde{H}_0$. Table 4 shows fuzzy p-values obtained by a one-sided test for the mean θ, $\tilde{H}_0 : \theta > \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \leq \tilde{\theta}_0$. Now that our fuzzy observations are crispier, the obtained fuzzy p-values are also crispier. This allows us to compute r^{acc} and r^{rej} that hold information. Indeed, on Table 3, which is about the one-sided test $\tilde{H}_0 : \theta < \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \geq \tilde{\theta}_0$, we can see with tests 1 and 2 that the fuzzy p-value strongly rejects the null hypothesis with a r^{rej} of 1. Tests 3, 4, 5 show us how different shapes for the null hypothesis yield to different non binary values r^{rej} and r^{acc}. At last, Table 3 also depicts how shifting null hypotheses from $(2.25, 2.35, 2.40)$ to $(2.5, 2.55, 2.6)$ gives values r^{acc} gradually going from 0 to 1 and values r^{rej} gradually going from 1 to 0. Table 4 illustrates the same results, but this time for the opposite side, $\tilde{H}_0 : \theta > \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \leq \tilde{\theta}_0$. Finally, for the two-sided test illustrated by Table 5, we again observe the same behaviour. Test 6 yields values $r^{acc} = 0$ and $r^{rej} = 1$. As opposed to it, test 1 gives values $r^{acc} = 1$ and $r^{rej} = 0$. Tests 2, 3, 4, 5 yield non binary values for r^{acc} and r^{rej}. As a whole, we observe that r^{acc} gradually goes from 1 to 0 while r^{rej} gradually goes from 0 to 1 when considering the tests 1, 2, 3, 4, 5, 6. This allows us to empirically discover the fuzzy region in between $(2.35, 2.4, 2.45)$ and $(2.45, 2.5, 2.5)$ where the null hypothesis is starting to be more and more rejected, respectively less and less rejected depending on the null hypothesis shape.

Figure 7: Crispier fuzzy p-value $\tilde{p} = (0, 0, 0.132)$. $r^{acc=0.39}$ and $r^{rej=0.61}$.



6 CONCLUSION

In this paper, we introduced a new procedure to find fuzzy p-values based on precedent works, which generalises the computation of crisp p-values. Our method revolves around generating a centred fuzzy bootstrapped statistic distribution to test and count how many of these bootstrapped observations are greater than an observed statistic. Then, we explained how the obtained fuzzy p-values could be interpreted as a ratio of a rejection or acceptance area over the area of the total area formed by the fuzzy p-value. We then enunciated a fuzzy hypothesis testing procedure to be able to compare fuzzy p-values to results obtained via this testing procedure. The main takeaway from this comparison is that fuzzy p-values tend to be very imprecise, with the observations getting fuzzier. However, it is still a helpful tool when the observations become crisper. Indeed, in the latter scenario, fuzzy p-values give us an idea of how much we're inside the fuzzy confidence interval or how much we're outside of it. In extreme cases, fuzzy p-values give the same binary results coming from crisp p-values.

REFERENCES

Arefi, M., Viertl, R., and Taheri, S. M. (2012). Fuzzy density estimation. *Metrika*, 75(1):5–22.

Berkachy, R. (2020). *The Signed Distance Measure in Fuzzy Statistical Analysis. Some theoretical, empirical and programming advances*. PhD thesis, University of Fribourg.

Berkachy, R. and Donzé, L. (2020). Fuzzy Confidence Intervals by the Likelihood Ratio with Bootstrapped Distribution. *Proceedings of the 12th International Joint Conference on Computational Intelligence*, pages 231–242.

Berkachy, R. and Donzé, L. (2017). Testing Fuzzy Hypotheses with Fuzzy Data and Defuzzification of the Fuzzy p-value by the Signed Distance Method. *Proceedings of the 9th International Joint Conference on Computational Intelligence*, pages 255–264.

Berkachy, R. and Donzé, L. (2022). Fuzzy Confidence Intervals by the Likelihood Ratio: Testing Equality of Means—Application on Swiss SILC Data. *SN Computer Science*, 3(5):374.

Chachi, J. and Taheri, S. M. (2011). Fuzzy confidence intervals for mean of Gaussian fuzzy random variables. *Expert Systems with Applications*, 38(5):5240–5244.

Grzegorzewski, P. and Romaniuk, M. (2021). Epistemic Bootstrap for Fuzzy Data. *Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)*, pages 538–545.

Hryniewicz, O. (2018). Statistical properties of the fuzzy p-value. *International Journal of Approximate Reasoning*, 93(J. Am. Stat. Assoc. 95 452 2000):544–560.

Kahraman, C., Otay, I., and Öztayşi, B. (2016). *Fuzzy Extensions of Confidence Intervals: Estimation for μ, σ^2, and p*, pages 129–154. Springer International Publishing, Cham.

Kahraman, C., Oztaysi, B., and Onar, S. C. (2019). Interval-Valued Intuitionistic Fuzzy Confidence Intervals. *Journal of Intelligent Systems*, 28(2):307–319.

Kruse, R. and Meyer, K. D. (1987). *Statistics with Vague Data*. D. Reidel Publishing Company.

Parchami, A. (2020). Fuzzy decision in testing hypotheses by fuzzy data: Two case studies. *Iranian Journal of Fuzzy Systems*.

Parchami, A., Taheri, S. M., and Mashinchi, M. (2018). Fuzzy p-value in testing fuzzy hypotheses with crisp data. *Springer Stat Papers*, 51.

Rosset, J. and Donzé, L. (2024a). New decision rules for fuzzy statistical inferences. In Ansari, J., Fuchs, S., Trutschnig, W., Lubiano, M. A., Gil, M. Á., Grzegorzewski, P., and Hryniewicz, O., editors, *Combining, Modelling and Analyzing Imprecision, Randomness and Dependence*, pages 405–412. Springer Nature Switzerland.

Rosset, J. and Donzé, L. (2024b). "fuzzy confidence intervals by bootstrapped fuzzy distributions". *IPMU 2024 LISBOA, 20th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (July 22–26, 2024) Lisboa*.

Shvedov, A. S. (2016). Quantile function of a fuzzy random variable and an expression for expectations. *Mathematical Notes*, 100(3-4):477–481.

Viertl, R. (2018). *Statistical Methods for Fuzzy Data*. John Wiley & Sons, Ltd.

Viertl, R. and Yeganeh, S. M. (2016). *Fuzzy Confidence Regions*, pages 119–127. Springer International Publishing, Cham.

Wu, H.-C. (2009). Statistical confidence intervals for fuzzy data. *Expert Systems with Applications*, 36(2):2670–2676.

Table 1: Fuzzy observations from Berkachy and Donzé (Berkachy and Donzé, 2020)

x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$	x_i	$\tilde{x}_i$
4	$(3,4,5)$	3	$(2,3,4)$	3	$(2,3,4)$	5	$(4,5,6)$	3	$(2,3,4)$
1	$(0,1,2)$	2	$(1,2,3)$	2	$(1,2,3)$	2	$(1,2,3)$	3	$(2,3,4)$

Note: $\bar{x} = 2.8$, $\bar{\tilde{x}} = (1.8, 2.8, 3.8)$.

Table 2: Fuzzy Hypothesis testing at a 5% significant level and fuzzy p-values.

Tests	$\tilde{H}_0$	L^{acc}	L^{rej}	Fuzzy p-value	r^{acc}	r^{rej}
1	(-0.3,-0.2,-0.1)	0	1	(0,0,0.032)	0	1
2	(0,0.5,0.8)	0	1	(0,0,1)	0.9025	0.0975
3	(0.8,0.85,0.9)	0	1	(0,0,1)	0.9025	0.0975
4	(0.95,1,1.05)	0.12	0.99	(0,0,1)	0.9025	0.0975
5	(1,1.5,2)	0.75	0.58	(0,0,1)	0.9025	0.0975
6	(1.8,1.9,2)	0.99	0.08	(0,0.012,1)	0.914	0.086
7	(1.7,1.75,1.78)	0.96	0.34	(0,0.01,1)	0.912	0.088
8	(2,2.5,3.5)	1	0	(0,0.5,1)	0.995	0.005
9	(2.4,2.45,2.5)	1	0	(0,0.32,1)	0.992	0.008
10	(2.6,2.8,3.2)	1	0	(0,0.87,1)	0.997	0.003
11	(0,3.8,6)	0.68	0.43	(0,0.08,1)	0.91	0.09
12	(0.5,4.5,8)	0.44	0.63	(0,0,0.132)	0.39	0.61
13	(7,8,9)	0	1	(0,0,0)	0	1

This Table shows first the levels of acceptance (L^{acc}) and rejection (L^{rej}) given by the fuzzy testing procedure of section 4. Then, the levels of acceptance (r^{acc}) and rejection (r^{rej}) of the associated fuzzy p-values are displayed. The fuzzy trapezoidal number $(0.933, 1.933, 3.6, 4.6)$ is the computed fuzzy confidence interval at a 95% confidence level for the distribution's mean. The values shown in red are in contradiction with the results of the testing procedure of section 4.

Table 3: Fuzzy statistics inferences at a 5% significance level. Fuzzy p-values for randomly generated fuzzy observations (one-sided test for the mean, θ, $\tilde{H}_0 : \theta < \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \geq \tilde{\theta}_0$).

Tests	$\tilde{H}_0$	Fuzzy p-value	r^{acc}	r^{rej}
1	(2.25,2.35,2.40)	(0,0.002,0.016)	0	1
2	(2.39,2.42,2.45)	(0.002,0.009,0.04)	0	1
3	(2.43,2.46,2.5)	(0.006,0.034,0.087)	0.32	0.68
4	(2.44,2.46,2.5)	(0.009,0.039,0.088)	0.44	0.56
5	(2.45,2.48,2.52)	(0.02,0.048,0.118)	0.68	0.32
6	(2.5,2.55,2.6)	(0.051,0.126,0.245)	1	0

This Table shows the levels of acceptance (r^{acc}) and rejection (r^{rej}) of different fuzzy null hypotheses via the obtained fuzzy p-values. The fuzzy mean of the dataset is $(2.693, 2.697, 2.704)$ and the associated fuzzy confidence interval at a 95% confidence level for the distribution's mean is $(2.42, 2.49, 2.94, 3.01)$.

Table 4: Fuzzy statistics inferences at a 5% significance level. Fuzzy p-values for randomly generated fuzzy observations (one-sided test for the mean, θ, $\tilde{H}_0 : \theta > \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \leq \tilde{\theta}_0$)).

Tests	$\tilde{H}_0$	Fuzzy p-value	r^{acc}	r^{rej}
1	(2.7,2.75,2.8)	(0.168,0.346,0.511)	1	0
2	(2.75,2.8,2.85)	(0.1,0.208,0.354)	1	0
3	(2.8,2.85,2.9)	(0.042,0.117,0.221)	0.993	0.007
4	(2.85,2.9,2.95)	(0.029,0.057,0.129)	0.85	0.15
5	(2.87,2.92,2.94)	(0.026,0.035,0.1)	0.52	0.48
6	(2.9,3,3.1)	(0.001,0.004,0.071)	0.1	0.9
7	(3,3.1,3.2)	(0,0,0.009)	0	1

This Table shows the levels of acceptance (r^{acc}) and rejection (r^{rej}) of different fuzzy null hypotheses via the obtained fuzzy p-values. The fuzzy mean of the dataset is $(2.693, 2.697, 2.704)$ and the associated fuzzy confidence interval at a 95% confidence level for the distribution's mean is $(2.42, 2.49, 2.94, 3.01)$.

Table 5: Fuzzy statistics inferences at a 5% significance level. Fuzzy p-values for randomly generated fuzzy observations (one-sided test for the mean, θ, $\tilde{H}_0 : \theta = \tilde{\theta}_0$ against $\tilde{H}_1 : \theta \neq \tilde{\theta}_0$).

Tests	$\tilde{H}_0$	Fuzzy p-value	r^{acc}	r^{rej}
1	(2.58,2.6,2.62)	(0.318,0.466,0.594)	1	0
2	(2.45,2.5,2.55)	(0.03,0.112,0.322)	0.984	0.016
3	(2.4,2.45,2.5)	(0.01,0.052,0.134)	0.7	0.3
4	(2.37,2.4,2.44)	(0.004,0.016,0.062)	0.06	0.94
5	(2.35,2.4,2.45)	(0.002,0.026,0.086)	0.74	0.26
6	(2.3,2.35,2.4)	(0,0.004,0.02)	0	1

This Table shows the levels of acceptance (r^{acc}) and rejection (r^{rej}) of different fuzzy null hypotheses through the obtained fuzzy p-values. The fuzzy mean of the dataset is $(2.693, 2.697, 2.704)$ and the associated fuzzy confidence interval at a 95% confidence level for the distribution's mean is $(2.42, 2.49, 2.94, 3.01)$.