Fuzzy Confidence Intervals by the Likelihood Ratio with Bootstrapped Distribution

Rédina Berkachy and Laurent Donzé
Applied Statistics and Modelling, Department of Informatics, Faculty of Management, Economics and Social Sciences, University of Fribourg, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland

Keywords: Bootstrap Technique, Likelihood Ratio, Fuzzy Confidence Interval, Fuzzy Statistics, Fuzzy Hypotheses, Fuzzy Data.

Abstract: We propose a complete practical procedure to construct a fuzzy confidence interval by the likelihood method where the observations and the hypotheses are considered to be fuzzy. We use the bootstrap technique to estimate the distribution of the likelihood ratio. For this step of the process, we mainly expose two algorithms: the first one consists on simply randomly drawing the bootstrap samples, and the second one is based on drawing observations by preserving the location and dispersion measures of the primary data set. This is achieved in accordance with a new metric written as $d_{SGD}^{\theta^*}$. It is built on the basis of the known signed distance measure. We also provide a simulation study to measure the performance of both bootstrap algorithms and their influence on the constructed confidence intervals. We illustrate our method via a numerical application where we construct fuzzy confidence intervals by the traditional and the defended methods. The aim is to highlight important differences between them.

1 INTRODUCTION AND MOTIVATION

A typical hypothesis testing procedure can be accomplished by, for example, constructing confidence intervals for a particular parameter. This method is widely used in practice. However, once we consider the data and/or the hypotheses to be fuzzy, the corresponding statistical methods have to be updated. Some approaches already exist in the theory of fuzzy sets. For instance, (Kruse and Meyer, 1987) presented a theoretical definition of fuzzy confidence intervals. Several researchers have afterwards proposed refined definitions of fuzzy confidence intervals. For instance, (Viertl and Yeganeh, 2016) proposed a definition of the so-called confidence regions. Their main application was in the Bayesian context. (Kahraman et al., 2016) described some approaches to the construction of fuzzy confidence intervals, as well as the concept of hesitant fuzzy confidence intervals. (Couso and Sanchez, 2011) provided an approach that considers the inner and outer approximations of confidence intervals in the context of fuzzy observations. Unfortunately, these various approaches are limited because they were all conceived to test a specific parameter with a pre-defined distribution. It would therefore be advantageous to develop a unified general approach to fuzzy confidence intervals.

In classical statistics, the likelihood ratio method is considered an alternative tool for the construction of confidence intervals. In the fuzzy environment, this method using uncertain data has multiple advantages. (Gil and Casals, 1988) used the likelihood ratio in a hypothesis testing procedure where fuzziness is contained in the data.

In (Berkachy and Donzé, 2019a), we proposed a practical procedure to construct confidence intervals by the likelihood ratio method which is seen in some sense general. The procedure can be easily adapted to specific cases. However, the distribution of the likelihood ratio is a priori unknown and has to be estimated or derived from strong assumptions. Under classical assumptions, we note that this ratio is known to be χ^2-distributed with degrees of freedom corresponding to the number of constraints applied to parameters. In this paper, we propose to use the bootstrap technique extended to the fuzzy environment to estimate the distribution of the likelihood ratio. A main contribution is to provide two algorithms to constitute the bootstrapped samples mainly using the location and dispersion characteristics calculated based on a new version of the signed distance measure written as

Berkachy, R. and Donzé, L.
Fuzzy Confidence Intervals by the Likelihood Ratio with Bootstrapped Distribution.
DOI: 10.5220/0010023602310242
In *Proceedings of the 12th International Joint Conference on Computational Intelligence (IJCCI 2020)*, pages 231-242
ISBN: 978-989-758-475-6

the $d_{SGD}^{\theta^\star}$ metric and detailed in (Berkachy, 2020). We highlight that the Expectation-Maximization (EM) algorithm based on the fuzziness of data described by (Denoeux, 2011) is used to calculate the maximum likelihood estimators (MLEs).

The defended procedure is considered efficient and computationally light because we do not have to consider every single value of the support set of the involved fuzzy numbers, as in the traditional fuzzy method. Indeed, four conveniently chosen values are used in the construction process. The presented calculations are done using the `R` package `FuzzySTs` shown in (Berkachy and Donzé, 2020).

The remainder of the paper proceeds as follows. In Section 2, we present the definition of the signed distance measure, followed by the definition of the $d_{SGD}^{\theta^\star}$ metric in Section 3. Section 4 is devoted to the construction of the traditional fuzzy confidence intervals. In Section 5, we discuss our concept of fuzzy confidence intervals constructed using the likelihood method and detail the two bootstrap algorithms to approximate the distribution of the likelihood ratio. In addition, a simulation study illustrates the proposed algorithms. We end the paper with Section 6 by presenting a numerical application where we estimate the traditional and the defended fuzzy confidence intervals and compare them.

2 THE SIGNED DISTANCE

The signed distance was used by (Yao and Wu, 2000) to rank fuzzy numbers. It has served in different contexts, such as the evaluation of linguistic questionnaires described in, for example, (Berkachy and Donzé, 2016) or hypotheses testing (see (Berkachy and Donzé, 2019b)). This distance has intrigued specialists because of its simplicity in terms of calculation and computation, and its directionality. The directionality of this distance means it can be negative or positive, indicating the direction between two fuzzy numbers. For instance, (Dubois and Prade, 1987) presented it as an expected value of a particular fuzzy number. It is defined as follows:

Definition 2.1 (Signed distance of a real value)**.** *The signed distance measured from the origin denoted by $d_0(a,0)$ for $a \in \mathbb{R}$ is a itself, that is $d_0(a,0) = a$.*

Definition 2.2 (Signed distance between two real values)**.** *The signed distance between two values a and $b \in \mathbb{R}$ is $d(a,b) = a - b$.*

Now, let $\tilde{X}$ and $\tilde{Y}$ be two sets of the class of fuzzy sets $\mathbb{F}(\mathbb{R})$. Their respective α-cuts are written as $\tilde{X}_\alpha$ and $\tilde{Y}_\alpha$ such that their left and right α-cuts denoted respectively by $\tilde{X}_\alpha^L$, $\tilde{X}_\alpha^R$, $\tilde{Y}_\alpha^L$ and $\tilde{Y}_\alpha^R$ are integrable for all $\alpha \in [0;1]$. We define the signed distance between the fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ as:

Definition 2.3 (Signed distance between two fuzzy sets)**.** *The signed distance d_{SGD} between $\tilde{X}$ and $\tilde{Y}$ is the mapping*

$$\begin{array}{rcl} d_{SGD} : \mathbb{F}(\mathbb{R}) \times \mathbb{F}(\mathbb{R}) & \to & \mathbb{R} \\ \tilde{X} \times \tilde{Y} & \mapsto & d_{SGD}(\tilde{X},\tilde{Y}), \end{array}$$

such that

$$\begin{aligned} d_{SGD}(\tilde{X},\tilde{Y}) &= \frac{1}{2}\int_0^1 \Big[\tilde{X}_\alpha^L(\alpha) + \tilde{X}_\alpha^R(\alpha) \\ &- \tilde{Y}_\alpha^L(\alpha) - \tilde{Y}_\alpha^R(\alpha)\Big] d\alpha. \qquad (1) \end{aligned}$$

The signed distance of a particular fuzzy number measured from the fuzzy origin $\tilde{0}$ is defined as:

Definition 2.4 (Signed distance of a fuzzy set)**.** *The signed distance of the fuzzy set $\tilde{X}$ measured from the fuzzy origin $\tilde{0}$ is given by:*

$$d_{SGD}(\tilde{X},\tilde{0}) = \frac{1}{2}\int_0^1 \Big[\tilde{X}_\alpha^L(\alpha) + \tilde{X}_\alpha^R(\alpha)\Big] d\alpha. \qquad (2)$$

3 THE $d_{SGD}^{\theta^\star}$ METRIC

Despite the simplicity and accessibility of the previously described distance d_{SGD}, it has some important drawbacks. First, it coincides with a central location measure. In other words, the effects of extreme values on a signed distance are strongly mitigated. Therefore, neither the shape of the fuzzy numbers nor the inner points between the extreme values affect this distance. Second, as detailed in (Berkachy, 2020), the signed distance cannot be defined as a full metric because it lacks topological characteristics, such as separability and symmetry. For all these reasons, we propose an L_2 metric denoted by $d_{SGD}^{\theta^\star}$, seen as a generalisation of the signed distance d_{SGD}. The metric $d_{SGD}^{\theta^\star}$ depends on a weight parameter called $\theta^\star$. With this new metric, we not only take into account the deviation in the shapes and its possible irregularities but also the central location measure. (Berkachy, 2020) proves that the measure $d_{SGD}^{\theta^\star}$ has the necessary and sufficient conditions to constitute a metric of fuzzy quantities.

We first define the so-called deviations of the shape of a particular fuzzy number written in terms of the distance d_{SGD} in the following way:

Definition 3.1. [(Berkachy, 2020)]. *Let $\tilde{X}$ be a fuzzy number with its α-level set $\tilde{X}_\alpha = [\tilde{X}_\alpha^L, \tilde{X}_\alpha^R]$ such that $\tilde{X}$*

$\in \mathbb{F}(\mathbb{R})$. *The left and right deviations of the shape of $\tilde{X}$ denoted by $dev^L\tilde{X}$ and $dev^R\tilde{X}$ can be given by:*

$$dev^L\tilde{X}(\alpha) = d_{SGD}(\tilde{X},\tilde{0}) - \tilde{X}^L_\alpha, \quad (3)$$
$$dev^R\tilde{X}(\alpha) = \tilde{X}^R_\alpha - d_{SGD}(\tilde{X},\tilde{0}), \quad (4)$$

where $d_{SGD}(\tilde{X},\tilde{0})$ is the signed distance of $\tilde{X}$ measured from the fuzzy origin $\tilde{0}$.

Now consider the following definition of the new metric $d^{\theta^\star}_{SGD}$ as expressed in (Berkachy, 2020).

Definition 3.2 (The $d^{\theta^\star}_{SGD}$ distance)**.** [(Berkachy, 2020)]. *Suppose two fuzzy numbers $\tilde{X}$ and $\tilde{Y}$ of the class of non-empty compact and bounded fuzzy numbers. Let $\theta^\star$ be the weight chosen for the modelling of the shape of these fuzzy numbers such that $0 \leq \theta^\star \leq 1$. Based on the signed distance between $\tilde{X}$ and $\tilde{Y}$, the L_2 metric $d^{\theta^\star}_{SGD}$ is the mapping*

$$\begin{array}{rcl} d^{\theta^\star}_{SGD} : \mathbb{F}(\mathbb{R}) \times \mathbb{F}(\mathbb{R}) & \rightarrow & \mathbb{R}^+ \\ \tilde{X} \times \tilde{Y} & \mapsto & d^{\theta^\star}_{SGD}(\tilde{X},\tilde{Y}), \end{array}$$

such that

$$d^{\theta^\star}_{SGD}(\tilde{X},\tilde{Y}) = \Big(\big(d_{SGD}(\tilde{X},\tilde{Y})\big)^2 + \theta^\star \Big(\int_0^1 \max\big(dev^R\tilde{Y}(\alpha) - dev^L\tilde{X}(\alpha), dev^R\tilde{X}(\alpha) - dev^L\tilde{Y}(\alpha)\big) d\alpha \Big)^2 \Big)^{\frac{1}{2}}. \quad (5)$$

It is useful to propose the concept of the nearest trapezoidal symmetrical fuzzy number. The intention is to show the direct relationship between the $d^{\theta^\star}_{SGD}$ metric and the signed distance measure. In the following context, the latter is considered as an optimum. This concept is defined by:

Definition 3.3 (Nearest trapezoidal fuzzy number)**.** [(Berkachy, 2020)]. *The nearest symmetrical trapezoidal fuzzy number $\tilde{S}$ written by the quadruple $\tilde{S} = [s_0 - 2\varepsilon, s_0 - \varepsilon, s_0 + \varepsilon, s_0 + 2\varepsilon]$ to a fuzzy number $\tilde{X}$ with respect to the metric $d^{\theta^\star}_{SGD}$ is given such that*

$$s_0 = d_{SGD}(\tilde{X},\tilde{0}), \quad (6)$$
$$\varepsilon = \frac{9}{14} d_{SGD}(\tilde{X},\tilde{0}) - \frac{3}{7}\int_0^1 \tilde{X}^L_\alpha (2-\alpha) d\alpha. \quad (7)$$

The proof can be found in (Berkachy, 2020). Note that this definition will be used in the upcoming sections to randomly generate samples with respect to the characteristics s_0 and ε.

4 TRADITIONAL FUZZY CONFIDENCE INTERVALS FOR A PRE-DEFINED PARAMETER

In an epistemic approach, the parameter θ for which the confidence interval is produced, is considered to be vague. Therefore, getting a fuzzy-type interval is a direct consequence of the fuzziness of the parameter. Fuzzy confidence intervals can be defined using, for example, the (Kruse and Meyer, 1987) approach, and many computation procedures can be derived from this definition. This includes the known approach based on considering a pre-defined distribution. Hereafter, we recall the definition and the construction procedure of a traditional fuzzy confidence interval.

Let us consider a random sample $X_1,\ldots,X_n$ of size n. Suppose this sample to be fuzzy. We denote by $\tilde{X}_1,\ldots,\tilde{X}_n$ its fuzzy perception. For a given parameter θ, we are interested in testing the following hypotheses:

$$H_0 : \theta = \theta_0 \quad \text{against} \quad H_1 : \theta \neq \theta_0.$$

One could construct a fuzzy confidence interval for θ to accomplish this task at a particular significance level denoted by δ. Based on a vague sample, we define a two-sided fuzzy confidence interval $\tilde{\Pi}$ as described in (Kruse and Meyer, 1987).

Definition 4.1 (Fuzzy confidence interval)**.** [(Kruse and Meyer, 1987)]. *We denote by $[\pi_1,\pi_2]$ a symmetrical confidence interval for a particular parameter θ at the significance level δ. A fuzzy confidence interval $\tilde{\Pi}$ is a convex and normal fuzzy set such that its left and right α-cuts, respectively written by $\tilde{\Pi}_\alpha = [\tilde{\Pi}^L_\alpha, \tilde{\Pi}^R_\alpha]$, are given as follows:*

$$\tilde{\Pi}^L_\alpha = \inf\big\{a \in \mathbb{R} : \exists x_i \in (\tilde{X}_i)_\alpha, \forall i = 1,\ldots,n, \text{ such that } \pi_1(x_1,\ldots,x_n) \leq a\big\}, \quad (8)$$
$$\tilde{\Pi}^R_\alpha = sup\big\{a \in \mathbb{R} : \exists x_i \in (\tilde{X}_i)_\alpha, \forall i = 1,\ldots,n, \text{ such that } \pi_2(x_1,\ldots,x_n) \geq a\big\}. \quad (9)$$

This fuzzy confidence interval is a $1-\delta$ confidence one if for a parameter θ, we have

$$P\big(\tilde{\Pi}^L_\alpha \leq \theta \leq \tilde{\Pi}^R_\alpha\big) \geq 1-\delta, \quad \forall \alpha \in [0;1]. \quad (10)$$

In the same way, we could also write a one-sided fuzzy confidence interval as:

Remark 4.1. *The α-level sets of a left one-sided fuzzy confidence interval at a confidence level $1-\delta$ denoted by $\tilde{\Pi}_\alpha$ are written as:*

$$\tilde{\Pi}_\alpha = [\tilde{\Pi}^L_\alpha, \infty],$$

and the α*-cuts of a right one-sided one are given by:*

$$\tilde{\Pi}_\alpha = [-\infty, \tilde{\Pi}^R_\alpha].$$

By way of example, it is useful to describe the two-sided fuzzy confidence interval for the mean related to the normal distribution. It can be written as follows:

Remark 4.2. *Let* $X_1, \dots, X_n$ *be a sample of size* n *drawn from a normal distribution with a known variance and* $\tilde{X}_1, \dots, \tilde{X}_n$ *be the corresponding fuzzy random variable.* $\overline{\tilde{X}}$ *is the fuzzy sample mean such that its left and right* α*-cuts are respectively written as* $(\overline{\tilde{X}})^L_\alpha$ *and* $(\overline{\tilde{X}})^R_\alpha$*. The two-sided fuzzy confidence interval for the mean of this fuzzy sample is written by its* α*-cuts in the following way:*

$$\begin{aligned} \tilde{\Pi}_\alpha &= [\tilde{\Pi}^L_\alpha, \tilde{\Pi}^R_\alpha] \\ &= \left[(\overline{\tilde{X}})^L_\alpha - u_{1-\frac{\delta}{2}}\frac{\sigma}{\sqrt{n}}, (\overline{\tilde{X}})^R_\alpha + u_{1-\frac{\delta}{2}}\frac{\sigma}{\sqrt{n}}\right], \end{aligned} \quad (11)$$

where σ *is the standard deviation and* $u_{1-\frac{\delta}{2}}$ *is the* $1-\frac{\delta}{2}$ *ordered quantile taken from the standard normal distribution.*

5 FUZZY CONFIDENCE INTERVALS BY THE LIKELIHOOD METHOD

Fuzzy confidence intervals suit the statistical inference very well. Following Definition 4.1, constructing a confidence interval for a particular parameter depends on a specific distribution. We therefore present a generalisation of the previous construction and give a practical tool to estimate a fuzzy confidence interval. In classical statistical theory, this task can be done using the so-called "likelihood ratio" method. For fuzzy contexts, we proposed in (Berkachy and Donzé, 2019a) a new approach to constructing fuzzy confidence intervals based on the concept of the likelihood ratio, conveniently considering the fuzziness contained in the variables. Note that the likelihood ratio has been used several times in fuzzy environments such as in (Gil and Casals, 1988) for hypotheses testing.

Let us recall the definition of the likelihood function in classical theory.

Definition 5.1 (Likelihood function)**.** *Consider* X_i*,* $i = 1, \dots, n$ *to be a sequence of random variables independent identically distributed (i.i.d). Let* x_i*,* $i = 1, \dots, n$ *be their corresponding realisations. We denote by* $f(x_i; \underline{\theta})$ *the probability density function (pdf) of the variable* X_i*. Consider* $\underline{\theta}$ *to be a vector of unknown parameters in the parameter space* Θ*. We define the likelihood function* $L(\underline{\theta}; x_i)$ *by:*

$$L(\underline{\theta}; x_i) = f(x_i; \underline{\theta}). \quad (12)$$

In this case, the expression $f(x_i; \underline{\theta})$ is called the likelihood function because it is now on a function of the vector of parameters $\underline{\theta}$ rather than x_i.

Now assume that the variable X_i is fuzzy, and consider its fuzzy perception. In other words, the fuzzy random variable (FRV) $\tilde{X}_i$ is such that its corresponding fuzzy realisation $\tilde{x}_i$ is associated with a measurable membership function written as $\mu_{\tilde{x}_i}$ in the sense of Borel, i.e. $\mu_{\tilde{x}_i} : x \to [0;1]$. Following the probability notions defined by (Zadeh, 1968), we could then expose the likelihood function described in the fuzzy context as:

Definition 5.2 (Likelihood function of a fuzzy observation)**.** *Consider* $\underline{\tilde{\theta}}$ *to be a vector of fuzzy parameters in the parameter space* Θ*. For a single fuzzy observation* $\tilde{x}_i$*, the likelihood function can be expressed by:*

$$L(\underline{\tilde{\theta}}; \tilde{x}_i) = P(\tilde{x}_i; \underline{\tilde{\theta}}) = \int_{\mathbb{R}} \mu_{\tilde{x}_i}(x) f(x; \underline{\tilde{\theta}}) dx. \quad (13)$$

This probability can also be expressed using the α-cuts of the involved fuzzy numbers.

Now we consider the fuzzy sample $\underline{\tilde{x}}$ composed of all the fuzzy realisations $\tilde{x}_i$ of the fuzzy random variable $\tilde{X}_i$. We can express the likelihood function $L(\underline{\tilde{\theta}}; \underline{\tilde{x}})$ by:

$$\begin{aligned} L(\underline{\tilde{\theta}}; \underline{\tilde{x}}) &= P(\underline{\tilde{x}}; \underline{\tilde{\theta}}) \\ &= \int_{\mathbb{R}} \mu_{\tilde{x}_1}(x) f(x; \underline{\tilde{\theta}}) dx \cdot \ldots \cdot \int_{\mathbb{R}} \mu_{\tilde{x}_n}(x) f(x; \underline{\tilde{\theta}}) dx \\ &= \prod_{i=1}^{n} \int_{\mathbb{R}} \mu_{\tilde{x}_i}(x) f(x; \underline{\tilde{\theta}}) dx. \end{aligned} \quad (14)$$

We conclude that the log-likelihood function written as $l(\underline{\tilde{\theta}}; \underline{\tilde{x}})$ can be given as follows:

$$\begin{aligned} l(\underline{\tilde{\theta}}; \underline{\tilde{x}}) &= \log L(\underline{\tilde{\theta}}; \underline{\tilde{x}}) \\ &= \log \int_{\mathbb{R}} \mu_{\tilde{x}_1}(x) f(x; \underline{\tilde{\theta}}) dx + \ldots \\ &\quad + \log \int_{\mathbb{R}} \mu_{\tilde{x}_n}(x) f(x; \underline{\tilde{\theta}}) dx. \end{aligned} \quad (15)$$

We call $\hat{\underline{\tilde{\theta}}}$ a maximum likelihood estimator (MLE) of the fuzzy parameter $\underline{\tilde{\theta}}$. The likelihood ratio is given by:

$$\frac{L(\underline{\tilde{\theta}}; \underline{\tilde{x}})}{L(\hat{\underline{\tilde{\theta}}}; \underline{\tilde{x}})},$$

such that $L(\underline{\tilde{\theta}}; \underline{\tilde{x}})$ is the likelihood function related to the fuzzy parameter $\underline{\tilde{\theta}}$, and $L(\hat{\underline{\tilde{\theta}}}; \underline{\tilde{x}})$ is the likelihood function depending on the estimator $\hat{\underline{\tilde{\theta}}}$ with $L(\hat{\underline{\tilde{\theta}}}; \underline{\tilde{x}}) \neq$

0 and finite. It is important in this case to write the logarithm of this ratio. This latter is nothing but the difference between the log-likelihood functions evaluated at $\hat{\tilde{\underline{\theta}}}$ and at $\tilde{\underline{\theta}}$. Let us then write the statistic LR given by:

$$LR = -2 \log \frac{L(\tilde{\underline{\theta}};\tilde{\underline{x}})}{L(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}})} = 2\left[l(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}}) - l(\tilde{\underline{\theta}};\tilde{\underline{x}})\right], \quad (16)$$

such that $L(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}}) \neq 0$, $L(\tilde{\underline{\theta}};\tilde{\underline{x}}) \neq 0$ and are both finite. Under classical statistical assumptions, the ratio LR is known to be asymptotically χ^2-distributed with a particular number of degrees of freedom. Fuzzy statistical theories do not have any clear indication if this asymptotical property can also be proved for the considered contexts. For this reason, in Section 5.2 we propose a methodology to solve this problem using bootstrap techniques.

Recall that our purpose in constructing a $100(1-\delta)\%$ confidence interval is to find every value of $\tilde{\theta}$ for which we reject or we do not reject the null hypothesis. To construct the required $100(1-\delta)\%$ fuzzy confidence interval, let η be the $(1-\delta)$-quantile of the distribution of LR. The confidence interval can then be given by:

$$2\left[l(\hat{\tilde{\theta}};\tilde{\underline{x}}) - l(\tilde{\theta};\tilde{\underline{x}})\right] \leq \eta. \quad (17)$$

It can equivalently be given by

$$l(\tilde{\theta};\tilde{\underline{x}}) \geq l(\hat{\tilde{\theta}};\tilde{\underline{x}}) - \frac{\eta}{2}. \quad (18)$$

This latter can be explained as the interval composed of all the possible values of $\tilde{\theta}$ for which the log-likelihood maximum varies by no more than $\frac{\eta}{2}$. We add that depending on LR, the constructed fuzzy confidence interval $\tilde{\Pi}_{LR}$ given by its left and right α-cuts $[(\tilde{\Pi}_{LR})^L_\alpha; (\tilde{\Pi}_{LR})^R_\alpha]$ has to insure the following equation

$$P\Big((\tilde{\Pi}_{LR})^L_\alpha \leq \theta \leq (\tilde{\Pi}_{LR})^R_\alpha\Big) \geq 1-\delta, \ \forall \alpha \in [0;1] \quad (19)$$

for every value of the parameter θ. Therefore, we propose the construction of fuzzy confidence intervals using the following procedure.

5.1 Procedure

Our idea is to revisit the methodology of constructing fuzzy confidence intervals using the likelihood ratio. In our case, the data set is considered to be imprecise. The log-likelihood becomes a function of fuzzy information. Recall that the parameter is considered to be fuzzy. It is then natural to see that the needed MLE estimator has to be fuzzy nature-based. Consequently, assume that the calculated crisp MLE estimator is modelled by a convenient fuzzy number. Accordingly, the support set of this fuzzy number is a set of crisp elements. Considering every element of this set in the calculation process of the log-likelihood function is computationally tedious. For this reason, we propose choosing specific values leading to the calculation of the so-called threshold points. The intersection between these threshold points and the log-likelihood curve will be particularly interesting for us in the process of calculating the fuzzy confidence interval.

To develop this idea, let us first expose the so-called standardising function. It is deliberately proposed to preserve the $[0;1]$-interval identity as a basic property of α-level sets. It is given by:

Definition 5.3 (Standardising function)**.** [(Berkachy, 2020)]. *Let $\tilde{\theta}$ be a fuzzy number with its membership function $\mu_{\tilde{\theta}}$ and $\theta \in$ supp $(\tilde{\theta})$. The standardising function I_{stand} is:*

$$\begin{aligned} I_{stand}: \quad & \mathbb{R} \ \rightarrow \ \mathbb{R} \\ & l(\theta,\tilde{\underline{x}}) \ \mapsto \ I_{stand}\big(l(\theta,\tilde{\underline{x}})\big) = \frac{l(\theta,\tilde{\underline{x}}) - I_a}{I_b - I_a}, \end{aligned}$$

where I_a and I_b are arbitrary real values such that $I_a \leq l(\theta,\tilde{\underline{x}}) \leq I_b$ and $I_a \neq I_b$. We have that $I_{stand}\big(l(\theta,\tilde{\underline{x}})\big)$ is bounded and $0 \leq I_{stand}(l(\theta,\tilde{\underline{x}})) \leq 1$.

The different steps of the calculation procedure can now be given as follows:

1. Consider a fuzzy parameter $\tilde{\theta}$. We first have to calculate the log-likelihood function $l(\tilde{\theta};\tilde{\underline{x}})$ described in Equation 16.
2. Next, from the support and the core sets defining the fuzzy number modelling the MLE estimator composed of an infinity of values, we choose the lower and upper bounds only. In this way, we reduce the number of considered elements to four, and we denote them by p, q, r and s, such that $p \leq q \leq r \leq s$. Consider supp$(\hat{\tilde{\theta}})$ and core$(\hat{\tilde{\theta}})$ to be the support and the core sets of $\hat{\tilde{\theta}}$, respectively. The four values p, q, r and s are given by:
$$p = \min(\text{supp}(\hat{\tilde{\theta}})); \quad q = \min(\text{core}(\hat{\tilde{\theta}})); \quad (20)$$
$$r = \max(\text{core}(\hat{\tilde{\theta}})) \text{ and } s = \max(\text{supp}(\hat{\tilde{\theta}})). \quad (21)$$
We know that the fuzzy parameter is bounded and the sets supp$(\hat{\tilde{\theta}})$ and core$(\hat{\tilde{\theta}})$ are not empty. It is then clear that the four values p, q, r and s always exist. We mention that this choice of elements is somehow evident specifically for the case of a symmetrical probability function because the left and right-hand sides of a log-likelihood function are monotonic and continuous.
3. Next, we estimate η. We propose to use the bootstrap technique developed in the next section.

4. Once the parameter η is estimated, we calculate the threshold values denoted by I_1, I_2, I_3 and I_4 corresponding respectively to the chosen values p, q, r and s. Thus, we affect θ by each of the four values on the right-hand side of Equation 19. They are then written in the following manner:

$$I_1 = l(p;\underset{\sim}{\tilde{x}}) - \frac{\eta}{2}; \quad I_2 = l(q;\underset{\sim}{\tilde{x}}) - \frac{\eta}{2}; \quad (22)$$

$$I_3 = l(r;\underset{\sim}{\tilde{x}}) - \frac{\eta}{2} \text{ and } I_4 = l(s;\underset{\sim}{\tilde{x}}) - \frac{\eta}{2}. \quad (23)$$

5. Next, we denote by $I_{\min}$ and $I_{\max}$ the minimum and maximum thresholds given by:

$$I_{\min} = \min(I_1, I_2, I_3, I_4), \quad (24)$$

$$\text{and } I_{\max} = \max(I_1, I_2, I_3, I_4). \quad (25)$$

It is important to find $I_{\min}$ and $I_{\max}$ and include them in the calculation process in order to cover the entirety of the interval of the possible values verifying Equation 19.

6. We are now interested in finding the intersection between the log-likelihood function and the threshold values I_1, I_2, I_3 and I_4. Let $\theta_1^{\star L}$, $\theta_2^{\star L}$, $\theta_3^{\star L}$, $\theta_4^{\star L}$ and $\theta_1^{\star R}$, $\theta_2^{\star R}$, $\theta_3^{\star R}$, $\theta_4^{\star R}$ be the intersection abscisses. The letters "L" and "R" refer to the left and right sides of a given entity. The abscisses can be calculated by solving the following equations:

$$l^L(\theta_1^{\star L};\underset{\sim}{\tilde{x}}) = I_1 \text{ and } l^R(\theta_1^{\star R};\underset{\sim}{\tilde{x}}) = I_1, \quad (26)$$

$$l^L(\theta_2^{\star L};\underset{\sim}{\tilde{x}}) = I_2 \text{ and } l^R(\theta_2^{\star R};\underset{\sim}{\tilde{x}}) = I_2, \quad (27)$$

$$l^L(\theta_3^{\star L};\underset{\sim}{\tilde{x}}) = I_3 \text{ and } l^R(\theta_3^{\star R};\underset{\sim}{\tilde{x}}) = I_3, \quad (28)$$

$$l^L(\theta_4^{\star L};\underset{\sim}{\tilde{x}}) = I_4 \text{ and } l^R(\theta_4^{\star R};\underset{\sim}{\tilde{x}}) = I_4. \quad (29)$$

7. We then find the minimum and maximum left intersection abscisses written as

$$\theta_{\inf}^{\star L} = \inf(\theta_1^{\star L}, \theta_2^{\star L}, \theta_3^{\star L}, \theta_4^{\star L}), \quad (30)$$

$$\text{and } \theta_{\sup}^{\star L} = \sup(\theta_1^{\star L}, \theta_2^{\star L}, \theta_3^{\star L}, \theta_4^{\star L}). \quad (31)$$

The minimum and maximum right intersection abscisses are analogously given by:

$$\theta_{\inf}^{\star R} = \inf(\theta_1^{\star R}, \theta_2^{\star R}, \theta_3^{\star R}, \theta_4^{\star R}), \quad (32)$$

$$\text{and } \theta_{\sup}^{\star R} = \sup(\theta_1^{\star R}, \theta_2^{\star R}, \theta_3^{\star R}, \theta_4^{\star R}). \quad (33)$$

Note that these left and right side intersection abscisses are single and real values.

8. These intersection abscisses and the previously calculated entities are consequently used to construct the α-cuts of the fuzzy confidence interval using the likelihood ratio method $\tilde{\Pi}_{LR}$. We propose to write the left and right α-cuts $(\tilde{\Pi}_{LR})_\alpha = \left[(\tilde{\Pi}_{LR})_\alpha^L; (\tilde{\Pi}_{LR})_\alpha^R\right]$ as follows:

$$(\tilde{\Pi}_{LR})_\alpha^L = \Big\{\theta \in \mathbb{R} \mid \theta_{\inf}^{\star L} \leq \theta \leq \theta_{\sup}^{\star L} \text{ and } \alpha = I_{stand}\big(l(\theta,\underset{\sim}{\tilde{x}})\big) = \frac{l(\theta,\underset{\sim}{\tilde{x}}) - I_{\min}}{I_{\max} - I_{\min}}\Big\}, \quad (34)$$

$$(\tilde{\Pi}_{LR})_\alpha^R = \Big\{\theta \in \mathbb{R} \mid \theta_{\inf}^{\star R} \leq \theta \leq \theta_{\sup}^{\star R} \text{ and } \alpha = I_{stand}\big(l(\theta,\underset{\sim}{\tilde{x}})\big) = \frac{l(\theta,\underset{\sim}{\tilde{x}}) - I_{\min}}{I_{\max} - I_{\min}}\Big\}. \quad (35)$$

We add that we are able to prove that the fuzzy confidence interval $\tilde{\Pi}_{LR}$ previously described verifies Definition 4.1. In addition, it can be proved that regarding the coverage rate, Equation 20 holds from a theoretical point of view. The complete proofs of both assumptions can be found in (Berkachy, 2020).

5.2 Bootstrap Technique for the Approximation of the Likelihood Ratio and Its Distribution

(Efron, 1979) formally proposed the bootstrap technique to empirically estimate a particular sampling distribution using some observed data. Based on a random primary sample drawn from an unknown distribution, his method seeks to draw a large number of samples and thus construct a so-called bootstrap distribution of the statistic of interest. This technique aims to ensure the estimation of such distributions using random-based procedures. It has also been thought in the fuzzy environment. For example, the bootstrap technique was used in the hypotheses testing procedure for the mean of fuzzy random variables as discussed in (Gonzalez-Rodriguez et al., 2006). (Montenegro et al., 2004) stated that a bootstrap methodology is considered to be computationally lighter than asymptotic designs, for example.

For our fuzzy context, our idea is to empirically estimate the distribution of the likelihood ratio LR shown in Equation 17—the difference of the log-likelihood function evaluated at $\hat{\underset{\sim}{\tilde{\theta}}}$ compared to the one evaluated at $\underset{\sim}{\tilde{\theta}}$. We propose the following two approaches to construct the bootstrap imprecise samples.

- The first approach is based on simply generating with replacement a number D of bootstrap samples. For each sample, we calculate the needed deviance. The corresponding algorithm is as follows:

Algorithm 1:

1. Consider a particular estimator $\hat{\tilde{\underline{\theta}}}$. Based on the primary fuzzy sample, compute the value of the deviance $2\left[l(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}}) - l(\tilde{\underline{\theta}};\tilde{\underline{x}})\right]$.
2. From the original data set, construct a bootstrap data set by drawing randomly with replacement a set of observations.
3. Calculate the bootstrapped deviance $2\left[l(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}}) - l(\tilde{\underline{\theta}};\tilde{\underline{x}})\right]^{\text{boot}}$.
4. Recursively repeat the Steps 2 and 3 a large number D of times. The aim is to construct the bootstrapped distribution composed of D values.
5. Find η, the $(1-\delta)$-quantile of the bootstrapped distribution of the *LR*.

- The second approach is to generate D samples by preserving the location and dispersion characteristics s_0 and ε, respectively, of the nearest symmetrical trapezoidal fuzzy numbers. These fuzzy numbers are calculated based on the primary data set as seen in Proposition 3.3. Algorithm 2 using the characteristics (s_0, ε) is given by:

Algorithm 2:

1. For each observation of the primary sample, calculate the set of characteristics (s_0, ε).
2. From the calculated set of characteristics (s_0, ε) related to the initial data set, randomly draw with replacement and with equal probabilities a set of characteristics (s_0, ε). Based on this set, construct a bootstrap sample.
3. For each bootstrap sample, calculate the deviance $2\left[l(\hat{\tilde{\underline{\theta}}};\tilde{\underline{x}}) - l(\tilde{\underline{\theta}};\tilde{\underline{x}})\right]^{\text{boot}}$.
4. Recursively repeat the Steps 2 and 3 a large number D of times. The aim is to construct the bootstrapped distribution composed of D values.
5. Find η, the $(1-\delta)$-quantile of the bootstrapped distribution of LR.

For our approach, it is crucial to calculate a maximum likelihood estimator. The fuzzy EM algorithm can be adopted as seen in (Denoeux, 2011). However, a drawback of this specific algorithm is that it produces a crisp estimator instead of a fuzzy one. Because we lack of methods for obtaining a fuzzy maximum likelihood estimator in such contexts, we propose modelling the calculated EM crisp-based estimator using a triangular fuzzy number. This crisp element will serve as the core of the required fuzzy number. Regarding its shape, we propose to use symmetrical triangles as a first step in reducing as much as possible the complexity that could be due to the choice of shapes. The `R` package `EM.Fuzzy` described in (Parchami, 2018) can be used to find the crisp estimators using the EM algorithm.

Finally, note that in practical settings the proposed detailed procedure and calculations can be easily computed using our `R` package `FuzzySTs` described in (Berkachy and Donzé, 2020) and developed for application purposes.

5.3 Simulation Study

Next, we propose a simulation study illustrating the use of the presented bootstrap algorithms in the process of calculating fuzzy confidence intervals. We randomly generate data sets from different characteristics and different sample sizes. Consider data sets composed of $N = 50$, 100 and 500 observations and taken from a normal distribution $N(5,1)$. To simplify the situation, we model the observations of our data sets by triangular symmetrical fuzzy numbers with a support set of spread 2.

We calculate the fuzzy confidence intervals using the likelihood ratio method for the theoretical mean of the generated data sets at the confidence level $1-\delta = 1-0.05$. Therefore, we have to estimate the bootstrapped quantile η. This is done for each data set using Algorithms 1 and 2 proposed in Section 5.2. In our previous studies, we remarked that the number of iterations did not really influence the outcome of the calculations. Therefore, we consider the case of $D = 1000$ iterations for all our calculations. However, the fuzzy EM algorithm for calculating EM estimators leads to crisp estimators instead of fuzzy ones. For this reason, we assume the following two ways of modelling the MLE estimator:

- the first way is by using a triangular symmetrical fuzzy number of spread 2;
- the second way is by using a triangular symmetrical fuzzy number of spread 1.

One of our interests is to investigate the influence of the intentionally chosen degree of fuzziness of these estimators on the characteristics of the constructed fuzzy confidence intervals. Note that we will additionally use the fuzzy sample mean as a fuzzy estimor for the sake of comparison.

Table 1 shows the 95%-quantiles of the bootstrapped distribution of the likelihood ratio LR for the cases of 50, 100 and 500 sample sizes. From this table, we can briefly remark that the quantiles depending on the sample sizes and/or the algorithm chosen are somehow close. Furthermore, it is clear that greater fuzziness, that is modelling the MLE estimator using a fuzzy number with spread 2, leads to a greater quantile compared to the case where the modelling fuzzy number is less fuzzy, that is, modelling the MLE estimator by a fuzzy number with spread 1.

Table 1: The 95%-quantiles of the bootstrapped distribution of LR - Case of a data set taken from a normal distribution $N(5,1)$ modelled using triangular symmetrical fuzzy numbers at 1000 iterations.

Algorithm 1			
Sample size	**N=50**	**N=100**	**N=500**
Bootstrap quantile using the sample mean	1.990	2.038	2.342
Bootstrap quantile using the MLE estimator (Spread 2)	1.809	1.927	2.181
Bootstrap quantile using the MLE estimator (Spread 1)	1.523	1.626	1.825
Algorithm 2			
Sample size	**N=50**	**N=100**	**N=500**
Bootstrap quantile using the sample mean	1.802	1.845	2.118
Bootstrap quantile using the MLE estimator (Spread 2)	1.854	1.971	2.201
Bootstrap quantile using the MLE estimator (Spread 1)	1.563	1.671	1.864

Based on the boostrapped quantiles shown in Table 1, we now calculate the fuzzy confidence intervals using the likelihood method following the instructions given in Section 5.1. Note that for the construction of confidence intervals, we will develop the case with $N = 500$ observations only. Table 2 gives the lower and upper bounds of the support and the core sets of the calculated fuzzy confidence intervals.

Let us first look at the influence of the choice of the bootstrap algorithm on the constructed confidence intervals. From Table 2, it is clear that no notable differences exist between the support and the core sets obtained by both algorithms. We can conclude that the choice of algorithms has no evident effect on the outcome of the approach. Therefore, although the design of both algorithms is different relative to points 1 and 2, similar results are depicted. Obviously, the small fluctuations in the bootstrapped quantiles given in Table 1 did not drastically influence the outcome of our approach.

Contrarily, regarding the fuzziness chosen for modelling the MLE estimator, we can clearly see a difference in the support sets of the calculated fuzzy confidence intervals. In fact, less fuzziness in the fuzzy number modelling the MLE estimator leads to a smaller support set of the obtained confidence interval. By way of example for Algorithm 1, the fuzzy confidence interval for the MLE estimator with spread 1 has a smaller support set (i.e. $[4.303; 5.685]$) than the fuzzy confidence interval for the MLE estimator with spread 2 (i.e. $[3.804; 6.184]$). Consequently, because the degree of fuzziness of the estimator directly affects the constructed fuzzy confidence interval, it is very important to carefully model this MLE estimator.

Next, we compare the constructed bootstrap fuzzy confidence intervals to the traditional fuzzy confidence interval $\tilde{\Pi}$ given by the trapezoidal fuzzy number $\tilde{\Pi} = (3.907, 4.907, 5.080, 6.080)$, as shown in Section 4. We can see that the bootstrap fuzzy confidence interval using the MLE estimators results in slightly larger core sets, while the characteristics of the obtained support sets differ between the cases depending on the degree of fuzziness of the MLE estimator. In this context, we add that once we use the fuzzy sample mean as an estimator in the calculation process, we get a fuzzy confidence interval for which the support and the core sets are tighter than the ones of the traditional fuzzy confidence interval $\tilde{\Pi}$.

Table 2: The fuzzy confidence interval by the likelihood ratio at the 95% significance level - Case of 500 observations taken from a normal distribution $N(5,1)$ modelled by triangular symmetrical fuzzy numbers.

Algorithm 1				
	Support set		**Core set**	
	Lower	**Upper**	**Lower**	**Upper**
fci using the sample mean	3.991	5.996	4.940	5.047
fci using the MLE estimator (Spread 2)	3.804	6.184	4.795	5.193
fci using the MLE estimator (Spread 1)	4.303	5.685	4.797	5.191
Algorithm 2				
	Lower	**Upper**	**Lower**	**Upper**
fci using the sample mean	3.991	5.996	4.945	5.042
fci using the MLE estimator (Spread 2)	3.803	6.184	4.795	5.193
fci using the MLE estimator (Spread 1)	4.303	5.685	4.797	5.191

Simulation Study on Coverage Rates. It is crucial to investigate the coverage rates corresponding to the fuzzy confidence intervals calculated using the defended likelihood method. We generated a large number of data sets composed of $N = 100$, 500 and 1000 observations. These data sets are considered to be uncertain. Every observation is modelled by a triangular symmetrical fuzzy number such that its spread is equal to 2. For these samples, we estimate fuzzy confidence intervals for the mean at the confidence level

$1-\delta = 1-0.05$. We are particularly interested in calculating the coverage rates of the constructed confidence intervals.

For this study, we used both Algorithms 1 and 2 to estimate the bootstrapped distribution of the likelihood ratio LR. These calculations were performed using the MLE estimators modelled by fuzzy numbers of spreads 1 and 2. For the sake of comparison, we also considered the fuzzy sample mean, similar to the previously described analysis. The fuzzy confidence intervals obtained using the likelihood method are then constructed and their coverage rates calculated. As a final step, we compare their coverage rates with the rates of the traditional fuzzy ones given in Equation 11.

It appears that the coverage rates of the bootstrap fuzzy confidence intervals calculated using Algorithms 1 and 2 are overall very close in the various setups. The difference is not noteworthy. Therefore, we propose elaborating the rates of the intervals using Algorithm 1 only.

Overall, the difference in the coverage rates of the bootstrap fuzzy intervals compared to the ones of the traditional fuzzy method is slight. This difference did not exceed 1.4% in all the cases. Using similar setups as in the previously described analysis, the results are developed as follows.

Regarding the data sets composed of 500 observations, we found that the coverage rate of the fuzzy confidence interval achieved by the likelihood method using the fuzzy sample mean is about 94.4% in the core set and about 100% in the support set. It is exactly the same as the rate for the traditional fuzzy confidence interval. Nevertheless, according to the fuzzy confidence intervals calculated using the spread 1 and spread 2 fuzzy numbers modelling the MLE estimators, the coverage rates are about 95.6% in the core set for both cases and about 100% in the support set. Remember that these rates are somehow acceptable in this context since theoretically a 95% confidence level has to be guaranteed. From these numbers, it is clear that the fuzziness contained in the MLE estimator is supposed to be uncertain. Although affecting its spread, the fuzziness did not actually influence the coverage rate of the calculated fuzzy confidence interval.

As a general interpretation of the outcome of our approach compared to the traditional fuzzy one, we can say that the LR fuzzy confidence interval using the MLE estimators modelled by fuzzy numbers seemed to be slightly less restrictive than the traditional fuzzy one. Therefore, once the fuzzy sample mean serves as an estimator, the support of the obtained interval appears to be smaller. In this context, it would be interesting to further investigate the behavior of the coverage rates in several setups and to find an appropriate theoretical method for calculating fuzzy MLE estimators.

6 NUMERICAL APPLICATION

Let us now discuss the construction of such confidence intervals where we suppose our data and the hypotheses to be fuzzy. We will consider the known normal distribution in order to simplify understanding the different steps.

Suppose a random sample composed of 10 observations $X_1,\dots,X_{10}$ given in Table 3. This sample is supposed to be taken from a normal distribution with a mean μ and a known variance $\sigma^2 = 1.29$, i.e. $N(\mu,\sigma^2 = 1.29)$. We consider this sample to be uncertain and model its fuzzy perception using triangular fuzzy numbers. The setups to estimate a particular fuzzy confidence interval for the mean μ at the confidence level $1-\delta = 1-0.05$ are given in the following steps:

Table 3: The data set and the fuzzified observations.

Index	X_i	Triangular Fuzzy Number
1	4	$(3,4,5)$
2	1	$(0,1,2)$
3	3	$(2,3,4)$
4	2	$(1,2,3)$
5	3	$(2,3,4)$
6	2	$(1,2,3)$
7	5	$(4,5,6)$
8	2	$(1,2,3)$
9	3	$(2,3,4)$
10	3	$(2,3,4)$
	$\overline{X} = 2.8$	$\overline{\tilde{X}} = (1.8,2.8,3.8)$

- **Model the Data:** Suppose the following modelling schema:
  - the value ”1” modelled by $\tilde{L}_1 = (0,1,2)$,
  - the value ”2” modelled by $\tilde{L}_2 = (1,2,3)$,
  - the value ”3” modelled by $\tilde{L}_3 = (2,3,4)$,
  - the value ”4” modelled by $\tilde{L}_4 = (3,4,5)$,
  - the value ”5” modelled by $\tilde{L}_5 = (4,5,6)$.

  Based on this, we obtain the modelled sample shown in Table 3.

- **Define the Required Test:** We test a fuzzy null hypothesis $\tilde{H}_0$ against a fuzzy alternative one $\tilde{H}_1$ for the mean μ at the significance level $\delta = 0.05$. Let us consider the following fuzzy hypotheses $\tilde{H}_0$ and $\tilde{H}_1$:
  $\tilde{H}_0^T = (1.8, 2, 2.3)$ against $\tilde{H}_1^T = (2.25, 2, 5)$.
- **Calculate the Fuzzy Sample Mean:** The fuzzy sample average $\overline{\tilde{X}}$ of the fuzzy perceptions of the $n = 10$ observations denoted by $\overline{\tilde{X}} = (1.8, 2.8, 3.8)$ can be written by its α-cuts $(\overline{\tilde{X}})_\alpha = [(\overline{\tilde{X}})^L_\alpha; (\overline{\tilde{X}})^R_\alpha] = [1.8 + \alpha; 3.8 - \alpha]$.

6.1 Estimation of the Traditional Fuzzy Confidence Interval for the Mean

A traditional fuzzy confidence interval for the mean μ at the confidence level $1-\delta = 1-0.05$ can be estimated as presented in Definition 4.1. We obtain the lower and upper bounds of the confidence interval at the confidence level $1-0.05$:

$$\begin{aligned}\tilde{\Pi}_\alpha &= \left[\tilde{\Pi}^L_\alpha; \tilde{\Pi}^R_\alpha\right] \\ &= \left[(\overline{\tilde{X}})^L_\alpha - u_{1-\frac{\delta}{2}}\frac{\sigma}{\sqrt{n}}; (\overline{\tilde{X}})^R_\alpha + u_{1-\frac{\delta}{2}}\frac{\sigma}{\sqrt{n}}\right] \\ &= \left[1.0965 + \alpha; 4.5034 - \alpha\right],\end{aligned}$$

where $u_{1-\frac{\delta}{2}} = u_{0.975} = 1.96$ is the 0.975-quantile of the normal distribution, $\sigma = 1.135$ is the standard deviation and $n = 10$ is the number of observations. The obtained interval is shown in Figure 1.

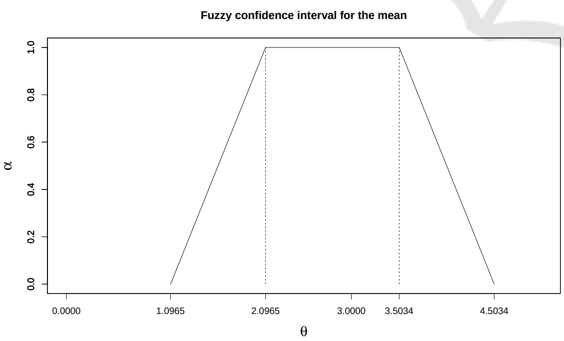


Figure 1: Traditional fuzzy confidence interval for the mean $\tilde{\Pi}$ - Section 6.

6.2 Estimation of the Fuzzy Confidence Interval for the Mean using the Likelihood Method

We now re-calculate the fuzzy confidence interval for the mean but use the likelihood methodology presented above. At the confidence level $1-\delta = 1-0.05$, the fuzzy confidence interval obtained using the likelihood method denoted by $\tilde{\Pi}_{LR}$ can then be constructed in the following way:

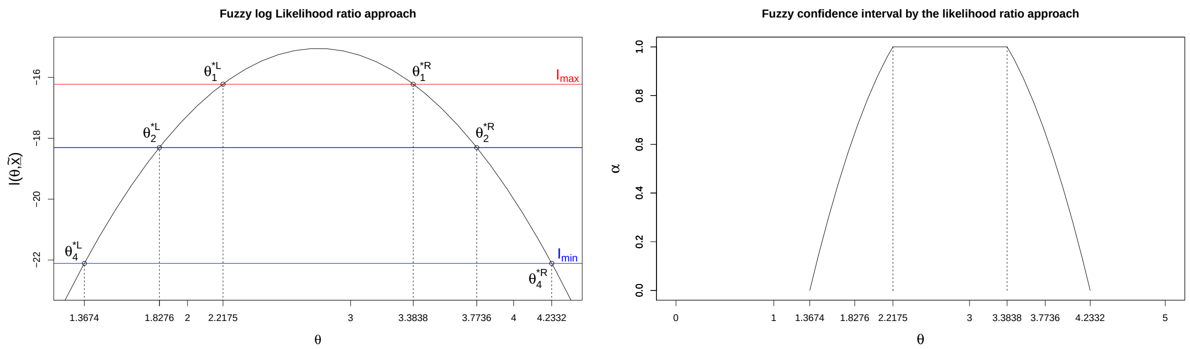
1. Consider the probability density function $f(x;\tilde{\theta})$ of the standard normal distribution such that $\sigma = 1.135$. For the fuzzy sample $\tilde{X}_i$, $i = 1,\dots,10$, we first calculate the log-likelihood function written as:

$$\begin{aligned} l(\tilde{\theta};\underset{\sim}{x}) &= \log\int_{\mathbb{R}}\mu_{\tilde{X}_1}(x)f(x;\tilde{\theta})dx + \dots \\ &\quad + \log\int_{\mathbb{R}}\mu_{\tilde{X}_{10}}(x)f(x;\tilde{\theta})dx \\ &= \log\int_3^4(3+x)f(x;\tilde{\theta})dx \\ &+ \log\int_4^5(5-x)f(x;\tilde{\theta})dx + \dots \\ &+ \log\int_2^3(2+x)f(x;\tilde{\theta})dx \\ &+ \log\int_3^4(4-x)f(x;\tilde{\theta})dx.\end{aligned}$$

2. Once we assume the sample to be fuzzy, we can consider the parameter to be fuzzy as well. However, we first calculate a crisp maximum likelihood estimator for the mean. The EM algorithm for the fuzzy context gives the crisp MLE estimator $\hat{\tilde{\theta}} = 3.6568$. Let us now model this crisp estimator using the following triangular symmetrical fuzzy number $(3.1568, 3.6568, 4.1568)$. We highlight that its support set is nothing but the interval $[3.1568; 4.1568]$, and its core set is reduced to the element 3.6568.
3. Let us now consider Algorithm 1 to estimate the distribution of the likelihood ratio LR by the bootstrap technique. At the significance level $\delta = 0.05$, the bootstrapped $(1-\delta)$-quantile η is estimated to be $\eta = 1.4778$. We then get $\frac{\eta}{2} = \frac{1.4778}{2} = 0.7389$. The threshold points I_1, I_2, I_3 and I_4 as described in Equations 23 and 24 have to be calculated afterwards. They are given by:

$$\begin{aligned} I_1 &= l(3.1568;\underset{\sim}{x}) - 0.7389 = -16.2258, \\ I_2 &= l(3.6568;\underset{\sim}{x}) - 0.7389 = -18.3079, \\ I_3 &= l(3.6568;\underset{\sim}{x}) - 0.7389 = -18.3079, \\ I_4 &= l(4.1568;\underset{\sim}{x}) - 0.7389 = -22.1101.\end{aligned}$$

   Note that the minimum and maximum thresholds shown in Figure 2(a) are $I_{\min} = \min(I_1, I_2, I_3, I_4) = -22.1101$, and $I_{\max} = \max(I_1, I_2, I_3, I_4) = -16.2258$.
4. The intersection points $\theta_1^{\star L}$, $\theta_2^{\star L}$, $\theta_3^{\star L}$, $\theta_4^{\star L}$ and $\theta_1^{\star R}$, $\theta_2^{\star R}$, $\theta_3^{\star R}$, $\theta_4^{\star R}$ have to be found as proposed in



(a) Fuzzy log-likelihood function for the mean and the intersection with the upper and lower bounds of the fuzzy parameter $\tilde{\bar{X}}$.

(b) Fuzzy confidence interval by likelihood ratio method $\tilde{\Pi}_{LR}$.

Figure 2: The construction process of the fuzzy confidence interval by the likelihood ratio method - Section 6.

Equations 27, 28, 29 and 30. We get:

$$\theta_1^{\star L} = 2.2175, \qquad \theta_1^{\star R} = 3.3838,$$
$$\theta_2^{\star L} = \theta_3^{\star L} = 1.8276, \qquad \theta_2^{\star R} = \theta_3^{\star R} = 3.7736,$$
$$\theta_4^{\star L} = 1.3674, \qquad \theta_4^{\star R} = 4.2332.$$

The minimum and maximum intersection abscisses $\theta_{\text{inf}}^{\star L}$, $\theta_{\text{sup}}^{\star L}$, $\theta_{\text{inf}}^{\star R}$ and $\theta_{\text{sup}}^{\star R}$ are given by:

$$\begin{aligned}
\theta_{\text{inf}}^{\star L} &= \inf(\theta_1^{\star L}, \theta_2^{\star L}, \theta_3^{\star L}, \theta_4^{\star L}) = 1.3674,\\
\theta_{\text{inf}}^{\star R} &= \inf(\theta_1^{\star R}, \theta_2^{\star R}, \theta_3^{\star R}, \theta_4^{\star R}) = 3.3838,\\
\theta_{\text{sup}}^{\star L} &= \sup(\theta_1^{\star L}, \theta_2^{\star L}, \theta_3^{\star L}, \theta_4^{\star L}) = 2.2175,\\
\theta_{\text{sup}}^{\star R} &= \sup(\theta_1^{\star R}, \theta_2^{\star R}, \theta_3^{\star R}, \theta_4^{\star R}) = 4.2332.
\end{aligned}$$

5. For the last step, we standardise the obtained function to the y-interval $[0;1]$ and get the following fuzzy confidence interval $\tilde{\Pi}_{LR}$ given by its left and right α-cuts $(\tilde{\Pi}_{LR})_\alpha = \left[(\tilde{\Pi}_{LR})_\alpha^L, (\tilde{\Pi}_{LR})_\alpha^R\right]$ as shown in Equations 35 and 36:

$$\begin{aligned}
(\tilde{\Pi}_{LR})_\alpha^L &= \Big\{\theta \in \mathbb{R} \mid 1.3674 \le \theta \le 2.2175\\
&\qquad \text{and } \alpha = \frac{l(\theta,\tilde{x}) + 22.1101}{5.8843}\Big\},\\
(\tilde{\Pi}_{LR})_\alpha^R &= \Big\{\theta \in \mathbb{R} \mid 3.3838 \le \theta \le 4.2332\\
&\qquad \text{and } \alpha = \frac{l(\theta,\tilde{x}) + 22.1101}{5.8843}\Big\}.
\end{aligned}$$

This latter is shown in Figure 2(b).

6.3 Comparison and Interpretation

At this stage, it would be interesting to compare the fuzzy confidence interval obtained using the likelihood approach $\tilde{\Pi}_{LR}$ to the traditional fuzzy confidence interval $\tilde{\Pi}$. If we superpose Figures 1 and 2(b), we can clearly see that the likelihood-based interval $\tilde{\Pi}_{LR}$ has a tighter support set than the traditional interval $\tilde{\Pi}$, i.e. $\tilde{\Pi}_{LR} \subseteq \tilde{\Pi}$. In the considered specific setups, we can say that our approach appears to be more restrictive. In other words, with our approach we tend to reject the hypothesis under study more often. We highlight that this interpretation can be slightly opposite in the situation with a larger support set of the fuzzy confidence interval. This could be mainly due to a greater amount of fuzziness contained in the fuzzy number modelling the MLE estimator.

Furthermore, concerning the shape of the obtained confidence intervals, i.e. their membership functions, it is clear that the shape of the traditional one strongly depends on the modelling procedure of the studied sample—in other words on the shape of the chosen fuzzy numbers. Note that this statement is similar in the case of the fuzzy sample mean. Contrarily, the fuzzy confidence interval obtained using the likelihood method relies directly on the probability density function connected to the considered data set. Finally, we can clearly see that the shape of the LR interval is more elaborated than the traditional one. This is considered an increase in the accuracy of such calculation methodologies.

7 CONCLUSION

This study proposed a complete procedure to estimate fuzzy confidence intervals using the likelihood ratio method. This required estimating the distribution of the likelihood ratio. A contribution of this research is the use of the bootstrap technique to accomplish this task. Two algorithms are discussed—a simple one and a more complex one based on preserving the location and dispersion measures related to a new metric called the $d_{SGD}^{\theta^\star}$ metric.

It is clear that such methodologies are computationally expensive. However, our procedure constitutes an affordable tool to reduce this computational complexity. Furthermore, our contribution is in some sense general, as it can be applied to a variety of estimators.

Finally, the main problem encountered is in the fuzziness contained in the fuzzy number modelling the maximum likelihood estimator. An investigation of different use cases depending on several modelling schemas for this estimator is welcome. Overall, a method of calculating a fuzzy-nature maximum likelihood estimator needs to be developed in future research.

REFERENCES

Berkachy, R. (2020). *The signed distance measure in fuzzy statistical analysis. Some theoretical, empirical and programming advances*. PhD thesis, University of Fribourg, Switzerland.

Berkachy, R. and Donzé, L. (2016). Individual and global assessments with signed distance defuzzification, and characteristics of the output distributions based on an empirical analysis. In *Proceedings of the 8th International Joint Conference on Computational Intelligence - Volume 1: FCTA,*, pages 75 – 82.

Berkachy, R. and Donzé, L. (2019a). Fuzzy confidence interval estimation by likelihood ratio. In *Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (EUSFLAT 2019)*.

Berkachy, R. and Donzé, L. (2019b). *Testing Hypotheses by Fuzzy Methods: A Comparison with the Classical Approach*, pages 1 – 22. Springer International Publishing, Cham.

Berkachy, R. and Donzé, L. (2020). FuzzySTs: Fuzzy statistical tools, R package, url = https://CRAN.R-project.org/package=FuzzySTs.

Chachi, J., Taheri, S. M., and Viertl, R. (2012). Testing statistical hypotheses based on fuzzy confidence intervals. *Austrian Journal of Statistics*, 41(4):267 –286.

Couso, I. and Sanchez, L. (2011). Inner and outer fuzzy approximations of confidence intervals. *Fuzzy Sets and Systems*, 184(1):68 – 83. Preference Modelling and Decision Analysis (Selected Papers from EUROFUSE 2009).

Denoeux, T. (2011). Maximum likelihood estimation from fuzzy data using the em algorithm. *Fuzzy Sets and Systems*, 183(1):72 – 91. Theme : Information processing.

Dubois, D. and Prade, H. (1987). The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24(3):279 – 300. Fuzzy Numbers.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Gil, M. A. and Casals, M. R. (1988). An operative extension of the likelihood ratio test from fuzzy data. *Statistical Papers*, 29(1):191 – 203.

Gonzalez-Rodriguez, G., Montenegro, M., Colubi, A., and Gil, M. A. (2006). Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems*, 157(19):2608 – 2613. Fuzzy Sets and Probability/Statistics Theories.

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.

Kruse, R. and Meyer, K. D. (1987). *Statistics with vague data*, volume 6. Springer Netherlands.

Montenegro, M., Colubi, A., Rosa Casals, M., and Ángeles Gil, M. (2004). Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika*, 59(1):31–49.

Parchami, A. (2018). EM.Fuzzy: Em algorithm for maximum likelihood estimation by non-precise information, R package, url = https://CRAN.R-project.org/package=EM.Fuzzy.

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

Yao, J.-S. and Wu, K. (2000). Ranking fuzzy numbers based on decomposition principle and signed distance. *Fuzzy sets and Systems*, 116(2):275 – 288.

Zadeh, L. (1968). Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications*, 23(2):421–427.