

# Flexible bootstrap based on the canonical representation of fuzzy numbers

Przemysław Grzegorzewski<sup>a,b</sup> and Olgierd Hryniewicz<sup>a</sup> and Maciej Romaniuk<sup>a,c</sup>

<sup>a</sup>Systems Research Institute, Polish Academy of Sciences,  
Newelska 6, 01-447 Warsaw, Poland

<sup>b</sup>Faculty of Mathematics and Information Science, Warsaw University of Technology,  
Koszykowa 75, 00-662 Warsaw, Poland

<sup>c</sup>Warsaw School of Information Technology, Newelska 6, 01-447 Warsaw, Poland  
{pgrzeg,hryniewi,mroman}@ibspan.waw.pl

## Abstract

A new resampling approach for simulating bootstrapped samples of fuzzy numbers is proposed. The secondary samples consist of fuzzy numbers which preserve the canonical representation (i.e., the value and ambiguity) of fuzzy numbers belonging to the primary sample, although may differ from the initial ones. This way the resulting bootstrap distribution has a richer support than obtained with the conventional method. Numerical experiments concerning two statistical tests for the expected value of a fuzzy random variable illustrate the suggested VA-bootstrap method and some of its properties.

**Keywords:** Bootstrap, Resampling, Canonical representation, Fuzzy data, Fuzzy numbers, Value, Ambiguity.

## 1 Introduction

The bootstrap, developed by Efron [7], is a popular resampling statistical technique for assessing uncertainty. Primary, it might be applied for estimating standard errors and computing confidence intervals. Generally, it appears extremely useful for estimating the distribution of a statistic of interest without using normal theory or any other parametric approach. Therefore, bootstrap methods become more and more popular in situations where inference involves complex procedures for which no theoretical results are available or remain questionable for the sample sizes met in practice.

A situation sketched above is typical in fuzzy modeling and statistical inference based on imprecise data expressed with fuzzy numbers. There, in most cases, the underlying distribution of the analysed phenomena remains unknown. Hence, in fuzzy context the bootstrap appears as not only helpful but sometimes even

the onliest tool that enables to conduct any statistical reasoning. In particular, the bootstrapped version of statistical tests for fuzzy data were considered, e.g., by Colubi et al. [3], Gil et al. [8], González-Rodríguez et al. [9, 10], Grzegorzewski and Ramos-Guajardo [12], Montenegro et al. [17] and Ramos-Guajardo and Lubiano [20]. Some other examples on the bootstrap application in fuzzy modeling of the real-life problems, like fuzzy rating in questionnaires [16], quality control in cheese manufacturing [19] or fuzzy nonparametric Shewhart's control chart construction [23] are worth mentioning too.

The classical bootstrap, in brief, involves drawing random samples with replacement from the initial sample of the experiment outcomes. As it is known, such bootstrap approach has an important disadvantage: it generates only such values which belong to the input (primary) sample. Consequently, nearly every bootstrap sample contains repeated values. Moreover, if the original sample size is small, all bootstrap samples consist of only few distinct values. It might be quite strange especially if the unknown original distribution is continuous. To overcome this inconvenience some modifications and improvements of the standard Efron's idea were proposed, like the balanced bootstrap by Davison et al. [4] or Graham et al. [11] or various kinds of the so-called smoothed bootstrap discussed, e.g., by Silverman and Young [22], Hall et al. [13] or De Angelis and Young [5].

The problem of excessive repetitions in bootstrap samples is also recognized as an unwanted situation in modeling complex phenomena with imprecise data. To increase the diversity of simulated results Romaniuk and Hryniewicz [21] proposed two resampling methods in which new triangular fuzzy numbers are generated from the primary sample by adding some incremental spreads for  $\alpha$ -cuts.

In this paper we propose a completely new resampling procedure based on the canonical representation of fuzzy numbers. The key idea is to generate trian-

gular or trapezoidal fuzzy numbers having the same canonical representation, i.e., the same value and the same ambiguity, as fuzzy observations in the primary sample. This way we may avoid any repetitions in the bootstrap sample since it consists of fuzzy numbers that can differ from the original observations. On the other hand they cannot differ too much since the main characteristics of the initial data are preserved. Indeed, by fixing the value and ambiguity of each original observation we determine its location and spread. Thus, although a membership function of the generated fuzzy number may differ a little bit from the membership function of the initial observation, the main information delivered by both fuzzy numbers is close.

Although one may ask why we restrict our attention to triangular and trapezoidal fuzzy numbers only, the reason is straightforward. As it has been noticed by many researchers the trapezoidal or triangular fuzzy numbers are most common in current applications mainly because they are easy to handle and have a natural interpretation. As noted by Trillas: “the problems that arise with vague predicates are less concerned with precision and are more of a qualitative type; thus they are generally written as linearly as possible. Normally it is sufficient to use a trapezoidal representation, as it makes it possible to define them with no more than four parameters” (see [14]). Moreover, even if the original data set consists of fuzzy numbers which are neither triangular nor trapezoidal, one may easily approximate them by such fuzzy numbers. In particular, an approximation algorithm which preserves the value and ambiguity of the original fuzzy number is given in [1], while the broad collection of approximation algorithms satisfying various requirements can be found in [2].

This paper is organized as follows. In Sec. 2 basic definitions and concepts related to fuzzy numbers and their representation is recalled. Next, the proposed resampling procedure for triangular and trapezoidal fuzzy numbers is described in Sec. 3 and Sec. 4, respectively. In both cases we provide the resampling algorithms that could be applied directly by the users. Finally, in Sec. 5 we deliver a few results of the simulation study performed to evaluate some properties of the suggested method. In particular, we compare the empirical size and the power of two tests equipped with the standard bootstrap procedure and our flexible bootstrap algorithm.

## 2 Fuzzy data

A **fuzzy number**  $A$  is a fuzzy set in  $\mathbb{R}$  which is normal, fuzzy-convex, has upper semicontinuous membership function  $A(x)$  and bounded support. A family of all fuzzy numbers will be denoted by  $\mathbb{F}(\mathbb{R})$ . An  $\alpha$ -cut of

a fuzzy number  $A$ , where  $\alpha \in [0, 1]$ , is defined by

$$A(\alpha) = \begin{cases} \{x \in \mathbb{R} : A(x) \geq \alpha\} & \text{if } \alpha \in (0, 1], \\ cl\{x \in \mathbb{R} : A(x) > 0\} & \text{if } \alpha = 0, \end{cases}$$

where  $cl$  stands for the closure operator. It is easily seen that the  $\alpha$ -cut  $A(\alpha)$  of a fuzzy number  $A$  is a closed interval  $A(\alpha) = [A_L(\alpha), A_U(\alpha)]$ .

The most often used fuzzy numbers are trapezoidal fuzzy numbers with membership functions of the form

$$A(x) = \begin{cases} \frac{x-a_1}{a_2-a_1} & \text{if } a_1 < x \leq a_2, \\ 1 & \text{if } a_2 \leq x \leq a_3, \\ \frac{a_4-x}{a_4-a_3} & \text{if } a_3 \leq x < a_4, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where  $a_1, a_2, a_3, a_4 \in \mathbb{R}$  such that  $a_1 \leq a_2 \leq a_3 \leq a_4$ . If  $a_2 = a_3$  then  $A$  is said to be a triangular fuzzy number.

Often, instead of declaring two points  $a_1$  and  $a_4$  describing the support of  $A$  and next two points for its core, i.e.,  $a_2$  and  $a_3$ , it is more convenient to use another parametrization through its location and a spread of its arm. Namely, let us define the following parameters

$$\begin{aligned} c &:= \frac{a_2 + a_3}{2}, & s &:= \frac{a_3 - a_2}{2}, \\ l &:= a_2 - a_1, & r &:= a_4 - a_3. \end{aligned} \quad (2)$$

One can easily identify  $c$  and  $s$  as the center and the half of the core, respectively, while  $l$  and  $r$  stand for the spread of the left and right arm of the membership function  $A(x)$ , respectively. Obviously,  $c \in \mathbb{R}$ , while  $s, l, r \geq 0$ . Using this notation a trapezoidal fuzzy number  $A$  would be denoted as  $A(c, s; l, r)$ . Similarly,  $A(c; l, r)$  stands for a triangular fuzzy number (since then  $s = 0$ ).

To simplify the representation of fuzzy numbers Delgado et al. [6] suggested two parameters value and ambiguity which represent some basic features of fuzzy numbers and hence they were called the *canonical representation of fuzzy numbers*.

A location of a fuzzy number  $A$  is characterized by its **value** defined as follows

$$\text{Val}(A) = \int_0^1 \alpha(A_U(\alpha) + A_L(\alpha))d\alpha, \quad (3)$$

whereas the **ambiguity** of  $A$ , given by

$$\text{Amb}(A) = \int_0^1 \alpha(A_U(\alpha) - A_L(\alpha))d\alpha, \quad (4)$$

is a measure of the global spread (or vagueness) of a fuzzy number  $A$ .

Since a value and ambiguity represent basic features of a fuzzy number, therefore two fuzzy numbers with the same ambiguity and value might be considered as similar (sometimes they are even treated as “almost equal”, see [6]).

One can easily find that the value and ambiguity of a trapezoidal fuzzy number  $A(c, s; l, r)$  are given as follows

$$\text{Val}(A) = c + \frac{r-l}{6}, \quad (5)$$

$$\text{Amb}(A) = s + \frac{r+l}{6}. \quad (6)$$

Obviously, if  $A(c; l, r)$  is a triangular fuzzy number, then its value is still given by (5), while its ambiguity reduces to

$$\text{Amb}(A) = \frac{r+l}{6}. \quad (7)$$

For more details on fuzzy numbers, their types, characteristics and approximations we refer the reader to [2].

### 3 Triangular observations

Let  $\tilde{X}_1, \dots, \tilde{X}_n$  denote a fuzzy random sample which realization is given by the triangular fuzzy numbers  $\tilde{x}_1, \dots, \tilde{x}_n \in \mathbb{F}(\mathbb{R})$ , where  $\tilde{x}_i = \tilde{x}_i(c_i; l_i, r_i)$ ,  $i = 1, \dots, n$ . In the following, this sample will be called the primary (initial) sample. Let us compute a value and ambiguity of each observation. This way we obtain a set of pairs

$$(\text{Val}(\tilde{x}_1), \text{Amb}(\tilde{x}_1)), \dots, (\text{Val}(\tilde{x}_n), \text{Amb}(\tilde{x}_n)), \quad (8)$$

where the corresponding values and ambiguities are calculated from (5) and (7), respectively.

The main idea of the proposed bootstrap technique is to generate randomly fuzzy observations from the set (8). Obviously, although the value and ambiguity characterize nicely a fuzzy number, they do not identify it completely. Of course, this imposes some restrictions, but we have still some room for the choice of the particular membership function. Let us consider how it works.

Given  $(\text{Val}^*, \text{Amb}^*)$  as an output of a random choice from (8), by (5) and (7) we obtain

$$\begin{cases} r-l = 6\text{Val}^* - 6c, \\ r+l = 6\text{Amb}^*, \end{cases}$$

moreover, by the definition,  $r, l \geq 0$ . Some immediate transformations yield in

$$\begin{cases} l = 3(\text{Amb}^* - \text{Val}^* + c), \\ r = 3(\text{Amb}^* + \text{Val}^* - c), \end{cases} \quad (9)$$

and hence, by  $r, l \geq 0$ , we obtain

$$\text{Val}^* - \text{Amb}^* \leq c \leq \text{Val}^* + \text{Amb}^*. \quad (10)$$

Now we are able to formulate the desired bootstrap algorithm. Keeping in mind (9) and (10), we obtain

#### Algorithm 1.

- 1) Given a fuzzy sample  $\tilde{x}_1, \dots, \tilde{x}_n \in \mathbb{F}(\mathbb{R})$ , where  $\tilde{x}_i = \tilde{x}_i(c_i; l_i, r_i)$ ,  $i = 1, \dots, n$ , compute the value  $\text{Val}(\tilde{x}_i)$  and ambiguity  $\text{Amb}(\tilde{x}_i)$  for each observation  $i = 1, \dots, n$ .
- 2) Let  $j := 1$ .
- 3) Generate randomly (with equal probabilities) a pair  $(\text{Val}_j^*, \text{Amb}_j^*)$  from  $(\text{Val}(\tilde{x}_1), \text{Amb}(\tilde{x}_1)), \dots, (\text{Val}(\tilde{x}_n), \text{Amb}(\tilde{x}_n))$ .
- 4) Generate a random value  $c_j^*$  from the uniform distribution on the interval  $[\text{Val}_j^* - \text{Amb}_j^*, \text{Val}_j^* + \text{Amb}_j^*]$ .
- 5) Compute  $l_j^* := 3[\text{Amb}_j^* - \text{Val}_j^* + c_j^*]$ .
- 6) Compute  $r_j^* := 3[\text{Amb}_j^* + \text{Val}_j^* - c_j^*]$ .
- 7) Let  $j := j + 1$ .
- 8) If  $j < n$  go to step 3.

Thus, following Algorithm 1 we receive the bootstrap sample of fuzzy triangular numbers  $\tilde{x}_1^*, \dots, \tilde{x}_n^*$ , where  $\tilde{x}_i^* = \tilde{x}_i^*(c_i^*; l_i^*, r_i^*)$ ,  $i = 1, \dots, n$ .

**Example 1.** Suppose that  $\tilde{x} = (6; 1, 2)$  is the randomly chosen observation from the initial sample. For this fuzzy number we have  $\text{Val}(\tilde{x}) = 6\frac{1}{6}$  and  $\text{Amb}(\tilde{x}) = \frac{1}{2}$ . Then, according to step 4 of Algorithm 1, the core  $c^*$  of the new bootstrapped fuzzy number is randomly generated from the uniform distribution on the interval  $[5\frac{2}{3}, 6\frac{2}{3}]$ . If, e.g.,  $c^* = 6\frac{1}{3}$  is selected, then by (9) we obtain  $l^* = 2$  and  $r^* = 1$ , so the resulting fuzzy number is  $\tilde{x}^* = (6\frac{1}{3}; 2, 1)$ . One may conclude immediately that although  $\tilde{x}^* \neq \tilde{x}$ , both fuzzy numbers have identical global location and spread, i.e.  $\text{Val}(\tilde{x}^*) = 6\frac{1}{6} = \text{Val}(\tilde{x})$  and  $\text{Amb}(\tilde{x}^*) = \frac{1}{2} = \text{Amb}(\tilde{x})$ .

### 4 Trapezoidal observations

Now suppose that the realization of a fuzzy random sample (i.e. the initial sample)  $\tilde{X}_1, \dots, \tilde{X}_n$  is given by the trapezoidal fuzzy numbers  $\tilde{x}_1, \dots, \tilde{x}_n \in \mathbb{F}(\mathbb{R})$ , where  $\tilde{x}_i = \tilde{x}_i(c_i, s_i; l_i, r_i)$ ,  $i = 1, \dots, n$ . As in Section 3 we compute the value and ambiguity of each observation and obtain the set

of pairs  $(\text{Val}(\tilde{x}_1), \text{Amb}(\tilde{x}_1)), \dots, (\text{Val}(\tilde{x}_n), \text{Amb}(\tilde{x}_n))$ , where the corresponding values and ambiguities are calculated from (5) and (6), respectively. Then, similarly as in the case of triangular data, we take randomly one pair and denote it by  $(\text{Val}^*, \text{Amb}^*)$ .

Given  $(\text{Val}^*, \text{Amb}^*)$ , by (5) and (6) we have

$$\begin{cases} r - l = 6\text{Val}^* - 6c, \\ r + l = 6\text{Amb}^* - 6s, \end{cases}$$

where  $s, r, l \geq 0$ , which is equivalent to

$$\begin{cases} l = 3(\text{Amb}^* - \text{Val}^* + c - s), \\ r = 3(\text{Amb}^* + \text{Val}^* - c - s). \end{cases} \quad (11)$$

Since  $r, l \geq 0$ , we obtain

$$\text{Val}^* - \text{Amb}^* + s \leq c \leq \text{Val}^* + \text{Amb}^* - s, \quad (12)$$

where  $s \geq 0$ . However, since the upper bound of (12) has to be not smaller than its lower bound, we obtain additionally that

$$0 \leq s \leq \text{Amb}^*. \quad (13)$$

Summing up the forementioned considerations and (11), (12) and (13), we obtain the following bootstrap algorithm for trapezoidal fuzzy data

#### Algorithm 2.

- 1) Given a fuzzy sample  $\tilde{x}_1, \dots, \tilde{x}_n \in \mathbb{F}(\mathbb{R})$ , where  $\tilde{x}_i = \tilde{x}_i(c_i, s_i; l_i, r_i)$ ,  $i = 1, \dots, n$ , compute the value  $(\text{Val}(\tilde{x}_i))$  and ambiguity  $\text{Amb}(\tilde{x}_i)$  for each observation  $i = 1, \dots, n$ .
- 2) Let  $j := 1$ .
- 3) Generate randomly (with equal probabilities) a pair  $(\text{Val}_j^*, \text{Amb}_j^*)$  from  $(\text{Val}(\tilde{x}_1), \text{Amb}(\tilde{x}_1)), \dots, (\text{Val}(\tilde{x}_n), \text{Amb}(\tilde{x}_n))$ .
- 4) Generate a random value  $s_j^*$  from the uniform distribution on the interval  $[0, \text{Amb}_j^*]$ .
- 5) Generate a random value  $c_j^*$  from the uniform distribution on the interval  $[\text{Val}_j^* - \text{Amb}_j^* + s_j^*, \text{Val}_j^* + \text{Amb}_j^* - s_j^*]$ .
- 6) Compute  $l_j^* := 3[\text{Amb}_j^* - \text{Val}_j^* + c_j^* - s_j^*]$ .
- 7) Compute  $r_j^* := 3[\text{Amb}_j^* + \text{Val}_j^* - c_j^* - s_j^*]$ .
- 8) Let  $j := j + 1$ .
- 9) If  $j < n$  go to step 3.

As the output of Algorithm 2 we receive the bootstrap sample of fuzzy trapezoidal numbers  $\tilde{x}_1^*, \dots, \tilde{x}_n^*$ , where  $\tilde{x}_i^* = \tilde{x}_i^*(c_i^*, s_i^*; l_i^*, r_i^*)$ ,  $i = 1, \dots, n$ .

**Example 2.** Let  $\tilde{x} = (6, 1; 2, 3)$  be an observation chosen randomly from the initial sample. For this fuzzy number we have  $\text{Val}(\tilde{x}) = 6\frac{1}{6}$  and  $\text{Amb}(\tilde{x}) = 1\frac{5}{6}$ . Then, according to step 4 of Algorithm 2 we generate a random value  $s^*$  from the uniform distribution on the interval  $[0, 1\frac{5}{6}]$ . If, e.g.,  $s^* = 1$  is selected, then in compliance with step 5, the value  $c^*$  is randomly chosen from the uniform distribution on the interval  $[5\frac{1}{3}, 7]$ . Suppose, that we have obtained  $c^* = 6\frac{2}{3}$ . Hence, by (11) we compute  $l^* = 4$  and  $r^* = 1$  and we finally obtain the new fuzzy number  $\tilde{x}^* = (6\frac{2}{3}, 1; 4, 2)$ . Obviously,  $\tilde{x}^* \neq \tilde{x}$ , however their fundamental characteristics are identical, i.e.  $\text{Val}(\tilde{x}^*) = 6\frac{1}{6} = \text{Val}(\tilde{x})$  and  $\text{Amb}(\tilde{x}^*) = 1\frac{5}{6} = \text{Amb}(\tilde{x})$ .

## 5 Simulation study

Nowadays, statistical tests for fuzzy sample attract attention of many researches (see, e.g., [8, 10, 12, 15, 17, 20]). Most of these tests utilize bootstrap methods to determine a null distribution under study. This is the reason that we also examine our new bootstrap methods suggested in Sec. 3 and Sec. 4 with two tests for the expected value. The first one is a bootstrapped version of the Körner test [15], while the second one is developed by Montenegro et al. [10, 17]. From now on, they will be denoted as the K-test and the M-test, respectively. Both tests were designed to verify a hypothesis on the expected value  $\mathbb{E}\tilde{X}$  of a random fuzzy variable, i.e.

$$H_0 : \mathbb{E}\tilde{X} = \tilde{v} \quad \text{vs.} \quad H_1 : \mathbb{E}\tilde{X} \neq \tilde{v}, \quad (14)$$

where  $\mathbb{E}\tilde{X}$  corresponds to the Aumann type mean of the fuzzy random variable  $\tilde{X}(\alpha)$  (see, [18]), and  $\tilde{v} \in \mathbb{F}(\mathbb{R})$  is a fixed fuzzy number considered further on as the true population mean.

Because of the limited space in this contribution we restrict our attention to triangular fuzzy numbers and, consequently, to the resampling method developed in Sec. 3 (further on denoted as VA-method). To be more specific, we consider three types of triangular fuzzy numbers to model the primary fuzzy random sample  $\tilde{X}_1, \dots, \tilde{X}_n$ . The first type (say, type A) is given by fuzzy numbers  $\tilde{X}_i(c_i; l_i, r_i)$  with centres (i.e., parameter  $c_i$ ) randomly generated from the standard normal distribution  $N(0, 1)$  and spreads (i.e.,  $l_i$  and  $r_i$ ) also randomly generated from the chi-square distribution with one degree of freedom  $\chi^2(1)$ . All random variables mentioned above are independent. A sample of the second type (say, type B) is generated as fuzzy

numbers  $\tilde{X}_i(c_i; l_i, r_i)$  with centres  $c_i$  from the standard normal distribution  $N(0, 1)$  and asymmetrical spreads, i.e.,  $l_i$  drawn from the exponential distribution  $\text{Exp}(2)$  with parameter 2 and  $r_i$  drawn from the exponential distribution  $\text{Exp}(4)$  with parameter 4. Fuzzy numbers of the third kind (type C) have centers generated from the standard normal distribution  $N(0, 1)$ , but both spreads are independently drawn from the uniform distribution  $U([0, 0.4])$ . Similar types of fuzzy numbers were considered in various studies, see [3, 15, 21].

For each type of fuzzy numbers, initial random samples of size  $n$  are used as input sets for the K-test and M-test. To compare various starting parameters we consider  $n = 5, 10, 30, 100$  corresponding to small and medium sample sizes. Then, each test is performed with the help of the classical bootstrap and the proposed VA-method. In both cases we consider several numbers  $b$  of the generated bootstrapped samples (namely,  $b = 100, 200, 1000$ ), which consist of  $n$  elements, to investigate a possible influence of  $b$  on the final results. In each experiment the whole re-sampling procedure is iterated 100000 times (see, e.g., [8, 10, 17, 20, 21] for more details concerning the similar comparisons).

$n$	5	10	30	100
$b$	100			
Boot.	0.16024	<b>0.10113</b>	<b>0.07006</b>	<b>0.06331</b>
VA	<b>0.15954</b>	0.10476	0.07616	0.06814
$b$	200			
Boot.	0.15438	<b>0.09686</b>	<b>0.0658</b>	<b>0.05795</b>
VA	<b>0.15236</b>	0.09833	0.0711	0.06207
$b$	1000			
Boot.	0.14834	<b>0.0911</b>	<b>0.06297</b>	<b>0.05449</b>
VA	<b>0.14523</b>	0.09421	0.06519	0.0585

Table 1: Empirical K-test size  $\hat{\alpha}$  for triangular numbers of type A.

As the essential benchmark we use the empirical size of the test  $\hat{\alpha}$  (i.e., the estimated percentage of the true null hypothesis rejections) and its relation to the nominal significance level  $\alpha$ . In our experiments we set the standard value  $\alpha = 0.05$ . Results given in Tables 1–6 show that both the classical approach (denoted by “Boot.”) and the suggested VA resampling procedure are very close to each other. It seems that the empirical sizes of the test  $\hat{\alpha}$  converge to the same limits as the initial sample  $n$  and the number of the bootstrap replications  $b$  increase. To ease noting the existing differences, those  $\hat{\alpha}$ , which are closer to the nominal significance level  $\alpha = 0.05$ , are given in boldface. For example, if  $n = 5, b = 100$  and  $\hat{\alpha} = 0.16024$ , as in Table 1, then the true null hypothesis was rejected in 16% of the cases (on average) for the small primary

sample (with 5 elements) and 100 replications of the bootstrapped sample.

$n$	5	10	30	100
$b$	100			
Boot.	0.17437	0.1078	0.07297	0.0633
VA	<b>0.168</b>	<b>0.10542</b>	<b>0.07108</b>	<b>0.05378</b>
$b$	200			
Boot.	0.16529	0.10371	0.06852	0.05739
VA	<b>0.16395</b>	<b>0.10009</b>	<b>0.06591</b>	<b>0.04933</b>
$b$	1000			
Boot.	0.16155	0.09762	0.06543	0.05515
VA	<b>0.15952</b>	<b>0.09628</b>	<b>0.06162</b>	<b>0.04674</b>

Table 2: Empirical K-test size  $\hat{\alpha}$  for triangular numbers of type B.

$n$	5	10	30	100
$b$	100			
Boot.	0.17553	0.10833	<b>0.07303</b>	<b>0.06285</b>
VA	<b>0.17183</b>	<b>0.10803</b>	0.07445	0.06319
$b$	200			
Boot.	<b>0.16669</b>	0.10309	<b>0.06878</b>	0.05775
VA	0.16806	<b>0.1021</b>	0.06895	<b>0.05746</b>
$b$	1000			
Boot.	<b>0.16255</b>	<b>0.0977</b>	0.06584	0.0556
VA	0.16352	0.09843	<b>0.06432</b>	<b>0.05432</b>

Table 3: Empirical K-test size  $\hat{\alpha}$  for triangular numbers of type C.

Going into details, when discussing the K-test (Tables 1–3), it seems that the VA-method dominates the classical bootstrap for type B fuzzy numbers. In the case of type A fuzzy numbers, the advantage of the VA-method appears only for small sample sizes. For type C fuzzy numbers situation is not completely clear, but it seems that the behavior of the VA-method improves and prevail the classical approach as the number of the bootstrap repetitions  $b$  increases.

In the case of the M-test (see Tables 4–6), the general comparison of the classical bootstrap and the VA-method shows that there is also no apparent winner: for small sample sizes sometimes the VA-method is better, otherwise the classical approach dominates. When we compare the situation under different number of bootstrap repetitions no obvious conclusion imposes.

Next step of the numerical experiment was the power study of the K-test and the M-test equipped with both the classical bootstrap or the VA-method. Because the null and the alternative hypotheses (14) in these tests concern the strict equality or inequality for a fuzzy number, the respective procedure has to be ap-



$n$	5	10	30	100
$b$	100			
Boot.	0.03375	<b>0.04906</b>	0.0562	0.05892
VA	<b>0.0339</b>	0.04347	<b>0.0489</b>	<b>0.05107</b>
$b$	200			
Boot.	<b>0.02988</b>	<b>0.04449</b>	<b>0.05196</b>	0.05473
VA	0.02915	0.03915	0.04493	<b>0.04597</b>
$b$	1000			
Boot.	<b>0.02748</b>	<b>0.04047</b>	<b>0.04916</b>	<b>0.05064</b>
VA	0.02603	0.03567	0.04078	0.0423

Table 4: Empirical M-test size  $\hat{\alpha}$  for triangular numbers of type A.

$n$	5	10	30	100
$b$	100			
Boot.	0.03634	<b>0.05033</b>	0.05777	0.05862
VA	<b>0.03724</b>	0.04912	<b>0.05488</b>	<b>0.05569</b>
$b$	200			
Boot.	0.03095	<b>0.04593</b>	0.05286	0.05344
VA	<b>0.03248</b>	0.04367	<b>0.04989</b>	<b>0.04911</b>
$b$	1000			
Boot.	0.02871	<b>0.04107</b>	<b>0.05009</b>	<b>0.05087</b>
VA	<b>0.02997</b>	0.04058	0.04564	0.0469

Table 5: Empirical M-test size  $\hat{\alpha}$  for triangular numbers of type B.

plied. To examine the power of the both tests, we estimated the number of rejections under increasing shift  $\epsilon \in \mathbb{R}$  of realizations of the initial fuzzy sample, when  $\epsilon = 0.25, 0.5, 0.75$ . We observed the behaviour of the procedures under study using the same values of parameters  $b$  and  $n$  as before. Because of lack of space we present to the reader the experimental results corresponding to fuzzy numbers of type C only (see Tables 7–8).

$n$	5	10	30	100
$b$	100			
Boot.	0.03246	0.04915	<b>0.05742</b>	<b>0.05851</b>
VA	<b>0.03352</b>	<b>0.05026</b>	0.0588	0.05903
$b$	200			
Boot.	0.02663	0.04431	<b>0.05325</b>	0.05327
VA	<b>0.02832</b>	<b>0.04433</b>	0.05351	<b>0.05314</b>
$b$	1000			
Boot.	0.02381	0.04006	<b>0.04954</b>	0.05113
VA	<b>0.02612</b>	<b>0.04026</b>	0.04874	<b>0.05046</b>

Table 6: Empirical M-test size  $\hat{\alpha}$  for triangular numbers of type C.

For the both tests the differences in the simulated power between the two considered approaches are rel-

atively small. Both resampling procedures behave in rather consistent way, i.e., the percentage of rejections generally grows as  $n$  or  $b$  increases. To emphasize, better results (if they exists) are given in boldface.

$n$	5	10	30	100
$\epsilon$	0.25			
$b$	100			
Boot.	<b>0.22496</b>	<b>0.199</b>	0.3177	0.72024
VA	0.22174	0.19778	<b>0.31772</b>	<b>0.72126</b>
$b$	200			
Boot.	0.21777	<b>0.19247</b>	<b>0.31134</b>	<b>0.71411</b>
VA	<b>0.21791</b>	0.19106	0.30949	0.71335
$b$	1000			
Boot.	0.21218	<b>0.18781</b>	0.30239	0.71188
VA	<b>0.21284</b>	0.18705	<b>0.30398</b>	<b>0.71282</b>
$\epsilon$	0.5			
$b$	100			
Boot.	<b>0.36195</b>	<b>0.43961</b>	<b>0.79944</b>	<b>0.9987</b>
VA	0.35853	0.43703	0.79904	0.99864
$b$	200			
Boot.	0.35157	0.42882	<b>0.79486</b>	0.99867
VA	<b>0.35244</b>	<b>0.42901</b>	0.79472	<b>0.99869</b>
$b$	1000			
Boot.	0.34748	<b>0.42479</b>	0.79131	0.99885
VA	<b>0.34767</b>	0.42217	<b>0.79273</b>	<b>0.99893</b>
$\epsilon$	0.75			
$b$	100			
Boot.	<b>0.54304</b>	<b>0.71307</b>	0.98386	1
VA	0.54079	0.7111	<b>0.98425</b>	1
$b$	200			
Boot.	0.53448	0.70511	<b>0.98386</b>	1
VA	<b>0.53467</b>	<b>0.70805</b>	0.9835	1
$b$	1000			
Boot.	<b>0.52919</b>	<b>0.70456</b>	<b>0.98447</b>	1
VA	0.52863	0.7022	0.98363	1

Table 7: K-test power analysis for triangular numbers of type C.

The differences between the bootstrap procedures are hardly discernible especially for the K-test, where it is difficult to designate the real winner. It can be seen also in Figure 1, where the percentages of rejections are shown for some range of smaller values of the shift (i.e.,  $\epsilon \in [0.05, 0.25]$ ) and  $n = 5, b = 100$ . The situation is slightly more clear in the case of the M-test. As it is seen in Table 8 and in Figure 2, the VA-method gives the higher percentages of rejections especially for the smallest sample size  $n$ , i.e.,  $n = 5$ .

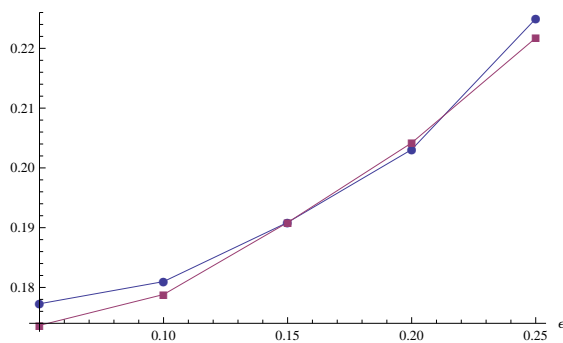


Figure 1: Power curves of the K-test for triangular numbers of type C for  $n = 5, b = 100$  (the classical bootstrap – circles, the VA-method – squares).

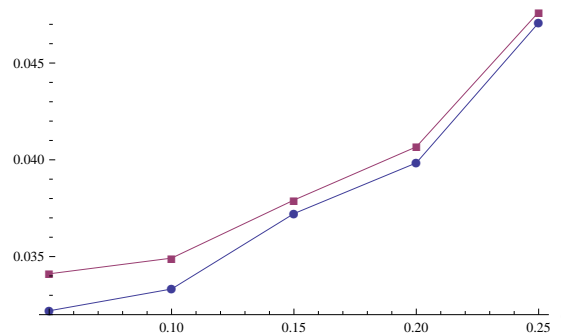


Figure 2: Power curves of the M-test for triangular numbers of type C for  $n = 5, b = 100$  (the classical bootstrap – circles, the VA-method – squares).

## 6 Conclusions

A flexible bootstrap approach for triangular and trapezoidal fuzzy numbers was proposed. Contrary to the classical Efron's method, our new algorithm generates values which are not necessary repetitions of the observations from the primary sample and are more diversified in some way. However, the resulting fuzzy numbers from the bootstrap samples preserve two important characteristics of the initial observations related to their canonical representation, i.e., the values and ambiguities of the original fuzzy data. Actually, there is no need to assume that the original sample consists of triangular or trapezoidal fuzzy numbers only. A declaration of a fuzzy number type is required just for generating bootstrap samples.

The paper implies two resampling algorithms ready for a direct use by the practitioners. The proposed methodology was also compared with the classical bootstrap via simulation study focused on two statistical tests. The initial results show that our new VA-method exhibits some promising properties and application potential, especially in the case of the small primary sample sizes and the K-test (for the true null hypothesis) or the M-test (when the power study is taken into account). Hence, more extended study would be still desirable. Moreover, further studies should go beyond hypothesis testing to examine the suggested approach in other fields of statistical inference. It seems also that the general idea of generating bootstrap samples which preserve some other characteristics of fuzzy numbers may lead to alternative flexible resampling procedures and it is worth a further study.

## References

- [1] A. Ban, A. Brândaș, L. Coroianu, C. Negruțiu, O. Nica, Approximations of fuzzy numbers by trapezoidal fuzzy numbers preserving the ambigu-

$n$	5	10	30	100
$\epsilon$	0.25			
$b$	100			
Boot.	0.04708	0.1045	<b>0.27674</b>	0.70835
VA	<b>0.04763</b>	<b>0.10472</b>	0.27601	<b>0.70968</b>
$b$	200			
Boot.	0.03834	<b>0.09749</b>	<b>0.26958</b>	0.70123
VA	<b>0.04177</b>	0.09739	0.26873	<b>0.70248</b>
$b$	1000			
Boot.	0.03571	<b>0.09049</b>	0.26048	0.69996
VA	<b>0.03792</b>	0.0899	<b>0.26096</b>	<b>0.7012</b>
$\epsilon$	0.5			
$b$	100			
Boot.	0.08898	<b>0.2769</b>	<b>0.76116</b>	<b>0.99863</b>
VA	<b>0.08967</b>	0.27502	0.76064	0.9985
$b$	200			
Boot.	0.07711	0.26214	<b>0.75629</b>	0.99851
VA	<b>0.08138</b>	<b>0.26332</b>	0.75503	<b>0.99855</b>
$b$	1000			
Boot.	0.07073	<b>0.25315</b>	0.75066	<b>0.99871</b>
VA	<b>0.07326</b>	0.25213	<b>0.75301</b>	0.9988
$\epsilon$	0.75			
$b$	100			
Boot.	0.15996	<b>0.52679</b>	0.97706	1
VA	<b>0.16177</b>	0.52473	<b>0.97759</b>	1
$b$	200			
Boot.	0.14084	0.51102	<b>0.97743</b>	1
VA	<b>0.14723</b>	<b>0.51362</b>	0.97705	1
$b$	1000			
Boot.	0.13156	<b>0.50551</b>	<b>0.97771</b>	1
VA	<b>0.13547</b>	0.50407	0.97696	1

Table 8: M-test power analysis for triangular numbers of type C.

- ity and value, *Computers and Mathematics with Applications* 61 (2011) 1379–1401.
- [2] A. Ban, L. Coroianu, P. Grzegorzewski, *Fuzzy Numbers: Approximations, Ranking and Applications*, Polish Academy of Sciences, Warsaw, 2015.
  - [3] A. Colubi, C. Fernández-García, M. Gil, Simulation of random fuzzy variables: An empirical approach to statistical/probabilistic studies with fuzzy experimental data, *IEEE Transactions on Fuzzy Systems* 10 (3) (2002) 384–390.
  - [4] A. C. Davison, D. V. Hinkley, E. Schechtman, Efficient bootstrap simulation, *Biometrika* 73 (1986) 555–566.
  - [5] D. De Angelis, G. A. Young, Smoothing the bootstrap, *International Statistical Review* 60 (1992) 45–56.
  - [6] M. Delgado, M. Vila, W. Voxman, On a canonical representation of a fuzzy number, *Fuzzy Sets and Systems* 93 (1998) 125–135.
  - [7] B. Efron, Bootstrap methods: Another look at the jackknife, *Annals of Statistics* 7 (1979) 1–26.
  - [8] M. Gil, M. Montenegro, G. González-Rodríguez, A. Colubi, M. Casals, Bootstrap approach to the multi-sample test of means with imprecise data, *Computational Statistics and Data Analysis* 51 (2006) 148–162.
  - [9] G. González-Rodríguez, A. Colubi, W. Trutschnig, Simulation of fuzzy random variables, *Information Sciences* 179 (5) (2009) 642–653.
  - [10] G. González-Rodríguez, M. Montenegro, A. Colubi, M. Gil, Bootstrap techniques and fuzzy random variables: Synergy in hypothesis testing with fuzzy data, *Fuzzy Sets and Systems* 157 (19) (2006) 2608–2613.
  - [11] R. Graham, D. V. Hinkley, P. W. M. John, S. Shi, Balanced design of bootstrap simulations, *Journal of the Royal Statistical Society, Series B* 52 (1990) 185–202.
  - [12] P. Grzegorzewski, A. Ramos-Guajardo, Similarity based one-sided tests for the expected value and interval data, in: J. M. Alonso, H. Bustince, M. Reformat (Eds.), *Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology*, Atlantis Press, 2015, pp. 960–966.
  - [13] D. T. J. Hall, P., J. P. Romano, On smoothing and the bootstrap, *Annals of Statistics* 17 (1989) 692–704.
  - [14] M. Jimenez, J. A. Rivas, Fuzzy number approximation, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 6 (1998) 68–78.
  - [15] R. Körner, An asymptotic  $\alpha$ -test for the expectation of random fuzzy variables, *Journal of Statistical Planning and Inference* 83 (2) (2000) 331–346.
  - [16] M. A. Lubiano, M. Montenegro, B. Sinova, S. De la Rosa de Saa, M. A. Gil, Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications, *European Journal of Operational Research* 251 (2016) 918–929.
  - [17] M. Montenegro, A. Colubi, M. Casals, M. Gil, Asymptotic and bootstrap techniques for testing the expected value of a fuzzy random variable, *Metrika* 59 (1) (2004) 31–49.
  - [18] M. Puri, D. Ralescu, Fuzzy random variables, *Journal of Mathematical Analysis and Applications* 114 (2) (1986) 409–422.
  - [19] A. Ramos-Guajardo, A. Blanco-Fernández, G. González-Rodríguez, Applying statistical methods with imprecise data to quality control in cheese manufacturing, in: P. Grzegorzewski, A. Kochanski, J. Kacprzyk (Eds.), *Soft Modeling in Industrial Manufacturing*, Springer, 2019, pp. 127–147.
  - [20] A. Ramos-Guajardo, M. Lubiano,  $k$ -sample tests for equality of variances of random fuzzy sets, *Computational Statistics and Data Analysis* 56 (4) (2012) 956–966.
  - [21] M. Romaniuk, O. Hryniewicz, Interval-based, nonparametric approach for resampling of fuzzy numbers, *Soft Computing* (in press), doi: 10.1007/s00500-018-3251-5.
  - [22] B. W. Silverman, G. A. Young, The bootstrap: To smooth or not to smooth?, *Biometrika* 74 (1987) 469–479.
  - [23] D. Wang, O. Hryniewicz, A fuzzy nonparametric Shewhart chart based on the bootstrap approach, *International Journal of Applied Mathematics and Computer Science* 25 (2) (2015) 389–401.