Two-Sample Dispersion Problem for Fuzzy Data

Przemyslaw Grzegorzewski[1,2(✉)]

[1] Systems Research Institute, Polish Academy of Sciences,
Newelska 6, 01-447 Warsaw, Poland
pgrzeg@ibspan.waw.pl
[2] Faculty of Mathematics and Information Science,
Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland

Abstract. The problem of comparing variability of two populations with fuzzy data is considered. A new permutation two-sample test for dispersion based on fuzzy random variables is proposed. A case-study illustrating the applicability of the suggested testing procedure is also presented.

Keywords: Fuzzy data · Fuzzy number · Fuzzy random variable · Permutation test · Test for dispersion · Test for scale

1 Introduction

Various two-sample statistical tests are designed to determine whether given two populations differ significantly. In such case we assume that the universe of discourse consists of two populations, say X and Y, with cumulative distribution functions F and G, respectively. Then, having a random sample of size n drawn from the X population and another random sample of size m drawn from the Y population, we consider the null hypothesis that these two samples are actually drawn from the same population, i.e. $H_0 : F = G$. One may verify H_0 against the general alternative hypothesis that the populations just differ in some way. The Kolmogorov-Smirnov test or the Wald-Wolfowitz run test are often used in this context (see e.g. [5]). However, they are really useful in preliminary studies only since affected by any type of difference between distributions, they are not very efficient in detecting any specific type of the difference like difference in location or difference in variablity. Other tests, like the Mann-Whitney-Wilcoxon test, the median test, etc. (see e.g. [5]) are particularly sensitive to differences in location when the populations are identical otherwise and hence cannot be expected to perform extremely well against other alternatives.

However, sometimes we need statistical procedures designed to detect differences in variability or dispersion instead of location. Indeed, comparison of

variability might be of interest in many areas including social sciences, biology, clinical trials, engineering, manufacturing and quality control, etc. Moreover, tests for the equality of variances are often required as a preliminary tool for the analysis of variance (ANOVA), dose–response modeling, discriminant analysis, etc.

It is important to emphasize that comparing variability is much harder than comparing measures of location. The famous F test assumes that both underlying populations are normally distributed and is not robust to departures from normality even asymptotically. Thus many nonparametric two-sample tests based on the ranks have been proposed for the scale problem. The best-known tests are the Ansari-Bradley test, the Mood test, the Siegel-Tukey test, the Klotz normal-scores test, the Sukhatme test, etc.

Designing tests for the dispersion problem turns out to be much more difficult in the case of imprecise or vague data which appear quite often in the real-life problems. In particular, human ratings based on opinions or associated with perceptions often lead to data that cannot be expressed in a numerical scale because they consist of intrinsically imprecise or fuzzy elements. Since they are also realizations of some random experiment, we are faced with random fuzzy structures that cannot be analyzed with classical statistical methods. Obviously, one may try to neglect, hide or remove imprecision but the most recommended approach is to consider it as a challenge for modeling and developing new inferential tools.

A general framework for such modeling is given by fuzzy random variables. However, besides mathematical elegence they also bring some fundamental difficulties. For instance, random fuzzy numbers are not linearly ordered so the aforementioned tests based on ranks cannot be directly applied in fuzzy environment. Depending on the context various test constructions have been proposed in the literature (for the overview we refer the reader e.g. to [7,8,11–14,16,18,19,21,26]). However, the dispersion problem with imprecise data has not beed considered very often. Ramos-Guajardo and Lubiano [26] proposed the bootstrap generalization of the Levene test for random fuzzy sets to examine homoscedasticity of k populations. Grzegorzewski [15] introduced two generalizations of the Sukhatme test for interval-valued data.

In this paper we suggest a permutation test for fuzzy data to compare variability of two populations. For motivations we turned back to the classical inference showing that permutation tests, like the bootstrap, require extremly limited assumptions. Indeed, permutation tests are totally distribution-free and require only exchangeability (i.e., under the null hypothesis we can exchange the labels on the observations without affecting the results). Classical permutation test are often more powerful than their bootstrap counterparts (see [9]). Permutation test are exact if all permutation are considered, while bootstrap tests are exact only for very large samples. Moreover, asymptotically permutation tests are usually as powerful as the most powerful parametric tests (see [1]). Keeping this in mind we combine the Pan test [22] and the Marozzi test [20] and then generalize them into the permutation testing procedure that handle fuzzy data.

The paper is organized as follows: in Sect. 2 we recall basic concepts related to fuzzy data modeling and operations on fuzzy numbers. Section 3 is devoted to fuzzy random variables. In Sect. 4 we introduce the two-sample test for the dispersion dedicated to fuzzy data. Next, in Sect. 5 we present some results of the simulation study and the case study with the proposed test. Finally, conclusions and some indications for the futher research are given in Sect. 6.

2   Fuzzy Data

A **fuzzy number** is an imprecise value characterized by a mapping $A : \mathbb{R} \to [0,1]$, called a membership function), such that its α-cut defined by

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

is a nonempty compact interval for each $\alpha \in [0,1]$. Operator cl in (1) denotes for the closure. Thus every fuzzy number is completely characterized both by its memberschip function $A(x)$ or by a family of its α-cuts $\{A_\alpha\}_{\alpha \in [0,1]}$. Two α-cuts are of special interest: $A_1 = \text{core}(A)$ known as the **core**, which contains all values which are fully compatible with the concept described by the fuzzy number A and $A_0 = \text{supp}(A)$ called the **support**, which are compatible to some extent with the concept modeled by A.

The most often used fuzzy numbers are **trapezoidal fuzzy numbers** (sometimes called *fuzzy intervals*) with membership functions of the form

$$A(x) = \begin{cases} \frac{x-a_1}{a_2-a_1} & \text{if } a_1 \leqslant x < a_2, \\ 1 & \text{if } a_2 \leqslant x \leqslant a_3, \\ \frac{a_4-x}{a_4-a_3} & \text{if } a_3 < x \leqslant a_4, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $a_1, a_2, a_3, a_4 \in \mathbb{R}$ such that $a_1 \leqslant a_2 \leqslant a_3 \leqslant a_4$. A trapezoidal fuzzy number (2) is often denoted as $\text{Tra}(a_1, a_2, a_3, a_4)$. Obviously, $a_1 = \inf \text{supp}(A)$, $a_2 = \inf \text{core}(A)$, $a_3 = \sup \text{core}(A)$ and $a_4 = \sup \text{supp}(A)$, which means that each trapezoidal fuzzy numbers is completely described by its support and core.

A fuzzy number A is said to be a **triangular fuzzy number** if $a_2 = a_3$, while if $a_1 = a_2$ and $a_3 = a_4$ we have the so-called **interval** (or rectangular) fuzzy number. The families of all fuzzy numbers, trapezoidal fuzzy numbers, triangular fuzzy number and interval fuzzy numbers will be denoted by $\mathbb{F}(\mathbb{R})$, $\mathbb{F}^T(\mathbb{R})$, $\mathbb{F}^\Delta(\mathbb{R})$ and $\mathbb{F}^I(\mathbb{R})$, respectively, where $\mathbb{F}^I(\mathbb{R}) \subset \mathbb{F}^\Delta(\mathbb{R}) \subset \mathbb{F}^T(\mathbb{R}) \subset \mathbb{F}(\mathbb{R})$.

Basic arithmetic operations in $\mathbb{F}(\mathbb{R})$ are defined through natural α-cut-wise operations on intervals. In particular, the sum of two fuzzy numbers A and B is given by the Minkowski addition of corresponding α-cuts, i.e.

$$(A+B)_\alpha = \left[\inf A_\alpha + \inf B_\alpha, \sup A_\alpha + \sup B_\alpha \right],$$

for all $\alpha \in [0,1]$. Similarly, the product of a fuzzy number A by a scalar $\theta \in \mathbb{R}$ is defined by the Minkowski scalar product for intervals, i.e. for all $\alpha \in [0,1]$

$$(\theta \cdot A)_\alpha = [\min\{\theta \inf A_\alpha, \theta \sup A_\alpha\}, \max\{\theta \inf A_\alpha, \theta \sup A_\alpha\}].$$

It is worth noting that a sum of trapezoidal fuzzy numbers is also a trapezoidal fuzzy number: if $A = \mathrm{Tra}(a_1, a_2, a_3, a_4)$ and $B = \mathrm{Tra}(b_1, b_2, b_3, b_4)$ then

$$A + B = \mathrm{Tra}(a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4). \tag{3}$$

Moreover, the product of a trapezoidal fuzzy number $A = \mathrm{Tra}(a_1, a_2, a_3, a_4)$ by a scalar θ is a trapezoidal fuzzy number

$$\theta \cdot A = \begin{cases} \mathrm{Tra}(\theta \cdot a_1, \theta \cdot a_2, \theta \cdot a_3, \theta \cdot a_4) & \text{if } \theta \geqslant 0, \\ \mathrm{Tra}(\theta \cdot a_4, \theta \cdot a_3, \theta \cdot a_2, \theta \cdot a_1) & \text{if } \theta < 0. \end{cases} \tag{4}$$

Unfortuntely, $(\mathbb{F}(\mathbb{R}), +, \cdot)$ has not linear but semilinear structure since in general $A + (-1 \cdot A) \neq \mathbb{1}_{\{0\}}$. Consequently, the Minkowski-based difference does not satisfy, in general, the addition/subtraction property that $(A + (-1 \cdot B)) + B = A$. To overcome this problem the so-called Hukuhara difference was defined as follows:

$$C := A -_H B \quad \text{if and only if} \quad B + C = A$$

Although now $A -_H A = \mathbb{1}_{\{0\}}$ or $(A -_H B) + B = A$ hold but the Hukuhara difference does not always exist. Therefore, one should be aware that subtraction in $\mathbb{F}(\mathbb{R})$ generally leads to critical problems and should be avoided, if possible.

At least some of the problems associated with the lack of a satisfying difference in constructing statistical tools for reasoning based on fuzzy observations could be overcome by using adequate metrics defined in $\mathbb{F}(\mathbb{R})$ – for the general overview see [2]. Obviously, one can define various metrics in $\mathbb{F}(\mathbb{R})$ but perhaps the most often used in statistical context is the one proposed by Gil et al. [6] and by Trutschnig et al. [27].

Let λ be a normalized measure associated with a continuous distribution having support in $[0,1]$ and let $\theta > 0$. Then for any $A, B \in \mathbb{F}(\mathbb{R})$ we define a metric D_θ^λ as follows

$$D_\theta^\lambda(A, B) = \left(\int_0^1 \left[(\mathrm{mid}\, A_\alpha - \mathrm{mid}\, B_\alpha)^2 + \theta \cdot (\mathrm{spr}\, A_\alpha - \mathrm{spr}\, B_\alpha)^2 \right] d\lambda(\alpha) \right)^{1/2}, \tag{5}$$

where $\mathrm{mid}\, A_\alpha = \frac{1}{2}(\inf A_\alpha + \sup A_\alpha)$ and $\mathrm{spr}\, A_\alpha = \frac{1}{2}(\sup A_\alpha - \inf A_\alpha)$ denote the mid-point and the radius of the α-cut A_α, respectively.

Both λ and θ correspond to some weighting: λ allows to weight the influence of each α-cut, while by a particular choice of θ one may weight the impact of the distance between the mid-points of the α-cuts (i.e. the deviation in location) in contrast to the distance between their spreads (i.e. the deviation in vagueness). In practice, the most common choice of λ is the Lebesgue measure on $[0,1]$), while

the most popular choice is $\theta = 1$ or $\theta = \frac{1}{3}$. It is worth noting that assuming $\theta = 1$ we obtain

$$D_1^\lambda(A, B) = \left(\int_0^1 \left[\frac{1}{2}(\inf A_\alpha - \inf B_\alpha)^2 + \frac{1}{2}(\sup A_\alpha - \sup B_\alpha)^2 \right] d\lambda(\alpha) \right)^{1/2}, \quad (6)$$

i.e. the metric which weights uniformly the two squared Euclidean distances and is equivalent to the distance considered in [4,10]. One may also notice that assuming $\theta = \frac{1}{3}$ we obtain

$$D_{1/3}^\lambda(A, B) = \sqrt{ \int_0^1 \left(\int_0^1 \left[A_\alpha^{[t]} - B_\alpha^{[t]} \right]^2 dt \right) d\lambda(\alpha) }, \quad (7)$$

where $A_\alpha^{[t]} = (1 - t)\inf A_\alpha + t \sup A_\alpha$, which means that $D_{1/3}^\lambda(A, B)$ aggregates uniformly the squared Euclidean distances between the convex combination of points in α-cuts representing A and B.

It should be stressed that whatever (λ, θ) is chosen D_θ^λ is an L^2-type metric in $\mathbb{F}(\mathbb{R})$ having some important and useful properties. It is translational invariant, i.e. $D_\theta^\lambda(A + C, B + C) = D_\theta^\lambda(A, B)$ for all $A, B, C, \in \mathbb{F}(\mathbb{R})$, and it is rotational invariant, i.e. $D_\theta^\lambda\big((-1) \cdot A, (-1) \cdot B\big) = D_\theta^\lambda(A, B)$ for all $A, B \in \mathbb{F}(\mathbb{R})$. Moreover, $(\mathbb{F}(\mathbb{R}), D_\theta^\lambda)$ is a separable metric space and for each fixed λ all D_θ^λ are topologically equivalent.

3   Fuzzy Random Variables

Suppose that the result of an experiment consists of random samples of imprecise data described by fuzzy numbers. To cope with such problem we need a model which grasps both aspects of uncertainty that appear in data, i.e. randomness (associated with data generation mechanism) and fuzziness (connected with data nature, i.e. their imprecision). To handle such data Puri and Ralescu [24] introduced the notion of a **fuzzy random variable** (also called a **random fuzzy number**).

Definition 1. *Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathbb{F}(\mathbb{R})$ is called a fuzzy random variable if for all $\alpha \in [0, 1]$ the α-cut function X_α is a compact random interval.*

In other words, X is a random fuzzy variable if and only if X is a Borel measurable function w.r.t. the Borel σ-field generated by the topology induced by D_θ^λ.

Puri and Ralescu [24] defined also the Aumann-type mean of a fuzzy random variable X as the fuzzy number $\mathcal{E}(X) \in \mathbb{F}(\mathbb{R})$ such that for each $\alpha \in [0, 1]$ the α-cut $\big(\mathcal{E}(X)\big)_\alpha$ is equal to the Aumann integral of X_α. It is seen that

$$\big(\mathcal{E}(X)\big)_\alpha = \big[\mathbb{E}(\mathrm{mid}\, X_\alpha) - \mathbb{E}(\mathrm{spr}\, X_\alpha), \mathbb{E}(\mathrm{mid}\, X_\alpha) + \mathbb{E}(\mathrm{spr}\, X_\alpha) \big].$$

To characterize dispersion of a fuzzy random variable X we can also define (see [17]) the D_θ^λ-Fréchet-type variance $\mathcal{V}(X)$, which is a nonnegative real number such that

$$
\mathcal{V}(X) = \mathbb{E}\Big(\big[D_\theta^\lambda(X, \mathcal{E}(X)) \big]^2 \Big)
$$
$$
= \int_0^1 \operatorname{Var}(\operatorname{mid} X_\alpha) d\lambda(\alpha) + \theta \int_0^1 \operatorname{Var}(\operatorname{spr} X_\alpha) d\lambda(\alpha).
$$

Given a sample of random fuzzy numbers $\mathbb{X} = (X_1, \ldots, X_n)$ a natural estimator of $\mathcal{E}(X)$ is the average $\overline{X} \in \mathbb{F}(\mathbb{R})$ such that for each $\alpha \in [0, 1]$

$$
\overline{X}_\alpha = \Big[\frac{1}{n} \sum_{i=1}^n \inf(X_i)_\alpha, \frac{1}{n} \sum_{i=1}^n \sup(X_i)_\alpha \Big] \tag{8}
$$
$$
= \Big[\frac{1}{n} \sum_{i=1}^n \operatorname{mid}(X_i)_\alpha - \frac{1}{n} \sum_{i=1}^n \operatorname{spr}(X_i)_\alpha, \frac{1}{n} \sum_{i=1}^n \operatorname{mid}(X_i)_\alpha + \frac{1}{n} \sum_{i=1}^n \operatorname{spr}(X_i)_\alpha \Big],
$$

while the estimator of $\mathcal{V}(X)$ is the D_θ^λ-type sample variance $S^2 \in \mathbb{R}$ given by

$$
S^2 = \frac{1}{n-1} \sum_{i=1}^n D_\theta^\lambda (X_i, \overline{X})^2. \tag{9}
$$

Although aforementioned constructions preserve many properties known from the real-valued inference, one should be aware of the problems typical of statistical reasoning with fuzzy data. As it was noted in Sect. 2, there are problems with subtraction of fuzzy numbers. Similar problems appear in the case of division of fuzzy numbers. Hence, it is advisable to avoid both operations wherever it is possible. Moreover, some difficulties in fuzzy data analysis is caused by the lack of universally accepted total ranking between fuzzy numbers. Another source of possible problems that appear in conjunction of randomness and fuzziness is the absence of suitable models for the distribution of fuzzy random variables. Even worse, there are not yet Central Limit Theorems for fuzzy random variables which can be applied directly in statistical inference.

The disadvantages mentioned above make the straightforward generalization of the classical statistical methodology into the fuzzy context either difficult or, sometimes, even impossible. For instance, in most cases we are not able to find the null distribution of a test statistic based on fuzzy data and, consequently, to find either the critical value or to compute the p-value required for rejection or acceptance of the hypothesis under study. To break through that problem some researchers propose to use the bootstrap [7,8,18,19,21,25,26].

In this paper we suggest another methodology based on permutations. For motivations we turn back to the classical inference which shows that permutation tests, like the bootstrap, require extremly limited assumptions. Bootstrap tests usually rely on assumption that successive observations are independent, while permutation tests require only exchangeability, i.e. under the null hypothesis we

can exchange the labels on the observations without affecting the results (obviously, if the observations in a sample are independent and identically distributed then they are exchangeable). In the real-valued framework one can also indicate two advantages of the permutation tests over the bootstrap tests. Firstly, permutation test are often more powerful than their bootstrap counterparts (see [9]). Secondly, permutation test are exact if all permutation are considered, while bootstrap tests are exact only for very large samples. Moreover, asymptotically permutation tests are usually as powerful as the most powerful parametric tests (see [1]). For more information on classical permutation tests we refer the reader to [9,23]. All these reasons indicate that the permutation test applied to fuzzy random variables might be also a competitive tool useful in statistical inference for imprecise data.

4   Permutation Test for Fuzzy Data to Compare Variability

Suppose, we observe independently two fuzzy random samples $\mathbb{X} = (X_1, \ldots, X_n)$ and $\mathbb{Y} = (Y_1, \ldots, Y_m)$ drawn from populations with unknown distributions function F and G, respectively. We want to verify the null hypothesis that both samples come from the same distribution, i.e.

$$H_0 : F(t) = G(t) \quad \text{for all } t \in \mathbb{R}, \tag{10}$$

against the alternative hypothesis that the dispersion of the distributions F and G differ (or against the one-sided alternative that the indicated distribution is more dispersed that the other one).

Most of the tests for scale assume that the distributions under study do not differ in location since possible location differences may mask differences in dispersion. Otherwise, the sample observations should be adjusted by subtrating the respective location parameters, like means or medians. If the true characteristics of location are not known we usually subtract their estimators.

Following remarks of Marozzi [20] on the resampling version of the Pan test [22] and the resampling framework for scale testing described by Boos and Brownie [3], we'll try to eliminate the location effects with sample means. However, keeping in mind problems with subtratiion in fuzzy environment described in Sect. 2, contrary to the crisp case, we do not consider the differences but the distances between sample observations and corresponding sample means calculated as in (8). Therefeore, further on instead of $\mathbb{X} = (X_1, \ldots, X_n)$ and $\mathbb{Y} = (Y_1, \ldots, Y_m)$ we consider the adjusted samples $\mathbb{V} = (V_1, \ldots, V_n)$ and $\mathbb{W} = (W_1, \ldots, W_m)$, respectively, where

$$V_i = D_\theta^\lambda(X_i, \overline{X}), \quad \text{for } i = 1, \ldots, n$$
$$W_j = D_\theta^\lambda(Y_j, \overline{Y}), \quad \text{for } j = 1, \ldots, m.$$

Now let us consider the following test statistics

$$T(\mathbb{X}, \mathbb{Y}) = \frac{\ln \overline{V} - \ln \overline{W}}{\sqrt{\frac{1}{n} \frac{S_V^2}{\overline{V}^2} + \frac{1}{m} \frac{S_W^2}{\overline{W}^2}}} \tag{11}$$

where S_V^2 and S_W^2 denote sample variances of $\mathbb{V}$ and $\mathbb{W}$, respectively, calculated by (9). Obviously, too big or too small values of (11) indicate that the null hypothesis should be rejected since the considered distributions differ in dispersion.

In the original Pan test [22] the decision whether to reject the null hypothesis is based on the test statistic valued with respect to some quantile from the t-Student distribution. However, Marozzi [20] showed that the resampling version of the Pan test should be rather preferred to the original one. In the case of fuzzy data any assumptions on the type of the underlying distribution of the samples are much more dubious than in the crisp case. For this reason we also consider here the permutation version of the Pan test. To carry out such a test we adapt the general idea of permutation tests to our fuzzy context.

The crucial idea of the proposed test construction is that the null hypothesis implies total exchangeability of observed data with respect to groups. Indeed, if H_0 holds then all available observations may be viewed as if they were randomly assigned to two groups but they come from the same population.

Let $\mathbb{Z} = \mathbb{X} \uplus \mathbb{Y}$, where $\uplus$ stands for the vector concatenation, so that the two samples are pooled into one, i.e. $Z_i = X_i$ if $1 \leqslant i \leqslant n$ and $Z_i = Y_{i-n}$ if $n + 1 \leqslant i \leqslant N$, where $N = n + m$.

Now, let $\mathbb{Z}^*$ denote a permutation of the initial dataset $\mathbb{Z}$. More formally, if $\nu = \{1, 2, \ldots, N\}$ and π_ν is a permutation of the integers ν, then $Z_i^* = Z_{\pi_\nu(i)}$ for $i = 1, \ldots, N$. Then the first n elements of $\mathbb{Z}^*$ is assigned to the first sample $\mathbb{Z}^*$ and the remaining m elements to $\mathbb{Z}^*$. In other words, it works like a random assignment of elements into two samples of the size n and m, respectively. Each permutation corresponds to some relabeling of the combined dataset $\mathbb{Z}$. Please, note that if H_0 holds then we are completely free to exchange the labels X or Y attributed to particular observations.

As a consequence of elements' exchangeability in $\mathbb{Z}^*$ under H_0 we can estimate the distribution of the test statistic T by considering all permutations of the initial dataset $\mathbb{Z}$ and computing a value of $T(\mathbb{Z}^*)$ corresponding to each permutation. Namely, given $\mathbb{Z} = \mathbb{z}$, where $\mathbb{z} = \mathbb{x} \uplus \mathbb{y}$, we take its permutation $\mathbb{z}^*$ and determine its adjustment with respect to sample means, i.e. we create two samples $\mathbb{v}^* = (v_1^*, \ldots, v_n^*)$ and $\mathbb{w}^* = (w_1^*, \ldots, w_m^*)$ as follows

$$v_i^* = D_\theta^\lambda\left(z_i^*, \frac{1}{n} \sum_{j=1}^n z_j^*\right), \quad \text{if } i = 1, \ldots, n$$

$$w_j^* = D_\theta^\lambda\left(z_i^*, \frac{1}{m} \sum_{i=j}^m z_j^*\right), \quad \text{if } i = n+1, \ldots, N.$$

Next, following (11) we compute its actual value corresponding to given permutation $\mathbb{z}^*$, i.e.

$$T(\mathbb{z}^*) = \frac{\ln \overline{v}^* - \ln \overline{w}^*}{\sqrt{\frac{1}{n}\frac{S_{v^*}^2}{(\overline{w}^*)^2} + \frac{1}{m}\frac{S_{w^*}^2}{(\overline{w}^*)^2}}}. \tag{12}$$

Finally, assuming K denotes a fixed number of drawings (usually not smaller than 1000), we calculate the p-value of our test. In the case on the one-sided upperer-tail test, i.e. when verifying $H_0 : F = G$ vs. H_1 stating that F is more disperded than G, we obtain

$$\text{p-value} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{1}\big(T(\mathbb{z}_k^*) \geqslant t_0\big), \tag{13}$$

where each $\mathbb{z}_k^* \in \mathcal{P}(\mathbb{z})$, $\mathbb{z}_k^* = \mathbb{x}_k^* \uplus \mathbb{y}_k^*$, and $t_0 = T(\mathbb{x}, \mathbb{y})$ stands for the test statistic value obtained for the original fuzzy samples $\mathbb{x}$ and $\mathbb{y}$.

For the one-sided lower-tail test, i.e. when verifying $H_0 : F = G$ vs. $H_1 : F$ is less disperded than G, we have

$$\text{p-value} = \frac{1}{K}\sum_{k=1}^{K}\mathbb{1}\big(T(\mathbb{z}_k^*) \leqslant t_0\big), \tag{14}$$

while for the two-sided test, i.e. when verifying $H_0 : F = G$ vs. $H_1 : F$ and G differ in dispersion, we obtain

$$\text{p-value} = \frac{1}{K}\left[\sum_{k=1}^{K}\mathbb{1}\big(T(\mathbb{z}_k^*) \geqslant |t_0|\big) + \mathbb{1}\big(T(\mathbb{z}_k^*) \leqslant -|t_0|\big)\right]. \tag{15}$$

5   Empirical Study

5.1   Simulations

We conducted some simulations to illustrate the behavior of the proposed test. To generate fuzzy samples from a trapezoidal-valued fuzzy random variable $X = \text{Tra}(\xi_1, \xi_2, \xi_3, \xi_4)$, where $\xi_1, \xi_2, \xi_3, \xi_4$ are real-valued random variables such that $\xi_1 \leqslant \xi_2 \leqslant \xi_3 \leqslant \xi_4$, the following characterization appears to be useful (see [19]): $c = \frac{1}{2}(\xi_3 + \xi_2) = \text{mid}_1 X$, $s = \frac{1}{2}(\xi_3 - \xi_2) = \text{spr}_1 X$, $l = \xi_2 - \xi_1$ and $r = \xi_4 - \xi_3$. Conversely, we have $\text{Tra}\langle c, s, l, r\rangle = \text{Tra}(c - s - l, c - s, c + s, c + s + r)$.

In our study we generated fuzzy observations $\mathbb{x} = (x_1, \dots, x_n)$ and $\mathbb{y} = (y_1, \dots, y_m)$ by simulating the four real-valued random variables $x_i = \langle c_{Xi}, s_{Xi}, l_{Xi}, r_{Xi}\rangle$ and $y_i = \langle c_{Yj}, s_{Yj}, l_{Yj}, r_{Yj}\rangle$, respectively, with the last three ones random variables in each quartet being nonnegative. In particular, we generated trapezoidal-valued fuzzy random variables using the following real-valued random variables: c_{Xi}, c_{Yj} from the normal distribution and $s_{Xi}, s_{Yj}, l_{Xi}, l_{Yj}, r_{Xi}$ and r_{Yj} from the uniform or chi-square distribution.
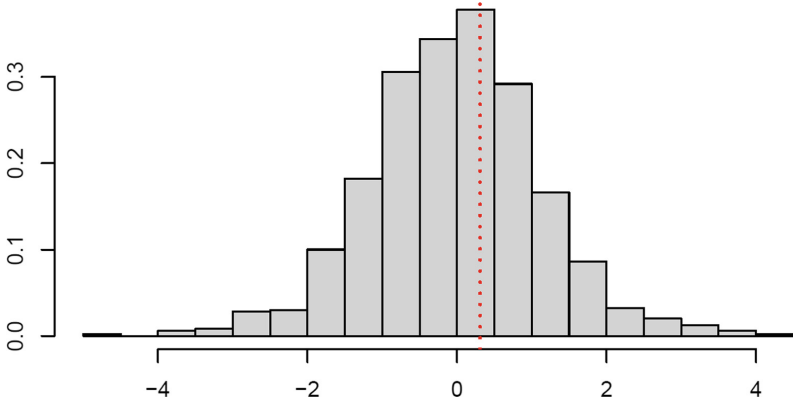
Fig. 1. Empirical null distribution of the permutation test with red vertical line indicatinges the value of the test statistic. (Color figure online)

An illustration how the test works, is shown in Fig. 1 and Fig. 2. Figure 1 shows a histogram made for the test statistic (11) null distribution obtained for two fuzzy samples of sizes $n = 10$ and $m = 12$. Both samples were generated as follows: c_X and c_Y came from the standard normal distribution $N(0, 1)$ and s_X, s_Y, l_X, l_Y and r_X, r_Y from the uniform distribution $U(0.0.5)$. In this case we have obtained $t_0 = 0.3088$, which is illustrated by a vertical line, while p-value = 0.384. A decision suggested by our test is: do not reject H_0.
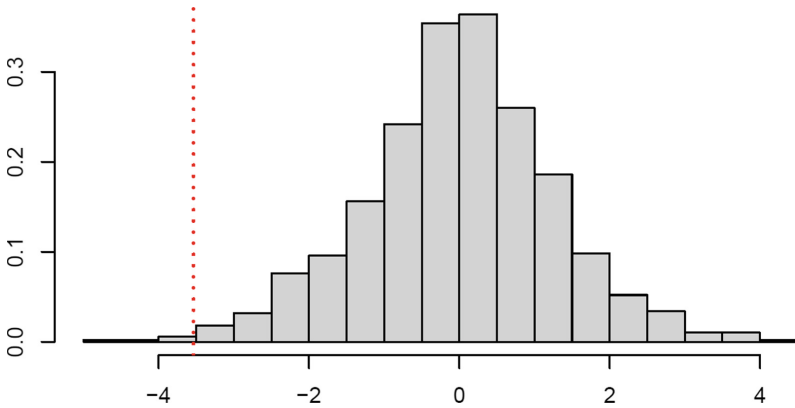

Fig. 2. Empirical null distribution of the permutation test with red vertical line indicatinges the value of the test statistic. (Color figure online)

On the other hand, in Fig. 2 we have a histogram made for the test statistic (11) null distribution obtained for two fuzzy samples of the same samle sizes as before but which differ in dispersion. Namely, c_X was generated from the

standard normal distribution $N(0, 1)$, but c_Y from $N(0, 2)$, while s_X, s_Y, l_X, l_Y and r_X, r_Y were, as befor, uniformly distributed from $U(0.0.5)$. In this case we have obtained $t_0 = -3.5373$, illustrated by a vertical line, and p-value $= 0.007$, leading to the decision: reject H_0.

Table 1. Empirical size of the test for various sample sizes.

n	m	empirical size	n	m	empirical size
10	10	0.021	10	15	0.024
20	20	0.024	10	20	0.015
50	50	0.023	10	50	0.009
100	100	0.034	50	100	0.026

We also examined the proposed permutation test with respect to its size. Therefore, 1000 simulations of the test performed on independent fuzzy samples comming from the same distribution were generated at the significance level 0.05. In each test $K = 1000$ permutations were drawn. Then empirical percentages of rejections under H_0 were determined. The results both for equal and nonequal sample sizes are gathered in Table 1. It is seen that our test is conservative. Moreover, this tendency deepens significantly as the imbalance of the sample sizes increases. These interesting results of the preliminary study of the proposed test properties indicate that further and more extensive study is highly recommended.

5.2 Case Study

Some statistical analyses of fuzzy data related to the Gamonedo cheese quality inspection was performed by Ramos-Guajardo and Lubiano [26] and Ramos-Guajardo et al. [25]. The Gamonedo cheese is a kind of a blue cheese produced Asturias, Spain. It experiences a smoked process and later on is let settle in natural caves or a dry place. To keep the quality of a cheese the experts (or tasters) usually express their subjective perceptions about different characteristics of the cheese, like visual parameters (shape, rind and appearance), texture parameters (hardness and crumbliness), olfactory-gustatory parameters (smell intensity, smell quality, flavour intensity, flavour quality and aftertaste) and their overall impression of the cheese.

Recently some of the tasters were proposed to express their subjective perceptions about the quality of the Gamonedo cheese by using trapezoidal fuzzy numbers. These fuzzy sets were determined in the following way: the set of values considered by the expert to be fully compatible with his/her opinion led to $\alpha = 1$-cut, while the set of values that he/she considered to be compatible with his/her opinion at some extent (i.e., the taster thought that it was not possible that the quality was out of this set) led to $\alpha = 0$-cut of a fuzzy number.

Then these two α-cuts were linearly interpolated to get the trapezoidal fuzzy set representing exppert's personal valuation. For more details on the data aquisition and analysis we refer the reader to Ramos-Guajardo et al. [25].

Table 2. Sample of the opinions of Expert 1 and 2 concerning the overall impression of the Gamonedo cheese (see [25])

Opinion	Expert 1	Expert 2	Opinion	Expert 1	Expert 2
1	$(65, 75, 85, 85)$	$(50, 50, 63, 75)$	21	$(65, 70, 76, 80)$	$(60, 64, 75, 85)$
2	$(35, 37, 44, 50)$	$(39, 47, 52, 60)$	22	$(75, 80, 86, 90)$	$(54, 56, 64, 75)$
3	$(66, 70, 75, 80)$	$(60, 70, 85, 90)$	23	$(65, 70, 73, 80)$	$(50, 50, 60, 66)$
4	$(70, 74, 80, 84)$	$(50, 56, 64, 74)$	24	$(70, 80, 84, 84)$	$(44, 46, 55, 57)$
5	$(65, 70, 75, 80)$	$(39, 45, 53, 57)$	25	$(55, 64, 70, 70)$	$(59, 63, 74, 80)$
6	$(45, 50, 57, 65)$	$(55, 60, 70, 76)$	26	$(64, 73, 80, 84)$	$(49, 50, 54, 58)$
7	$(60, 66, 70, 75)$	$(50, 50, 57, 67)$	27	$(50, 56, 64, 70)$	$(55, 60, 70, 75)$
8	$(65, 65, 70, 76)$	$(65, 67, 80, 87)$	28	$(55, 55, 60, 70)$	$(44, 47, 53, 60)$
9	$(60, 65, 75, 80)$	$(50, 50, 65, 75)$	29	$(60, 70, 75, 80)$	$(19, 20, 30, 41)$
10	$(55, 60, 66, 70)$	$(50, 55, 64, 70)$	30	$(64, 71, 80, 80)$	$(40, 44, 50, 60)$
11	$(60, 65, 70, 74)$	$(39, 46, 53, 56)$	31	$(50, 50, 55, 65)$	$(50, 50, 59, 66)$
12	$(30, 46, 44, 54)$	$(19, 29, 41, 50)$	32	$(50, 54, 60, 65)$	$(50, 53, 60, 66)$
13	$(60, 65, 75, 75)$	$(40, 47, 52, 56)$	33	$(65, 75, 80, 86)$	$(50, 52, 58, 61)$
14	$(70, 75, 85, 85)$	$(54, 55, 65, 76)$	34	$(50, 55, 60, 66)$	$(60, 65, 72, 80)$
15	$(44, 45, 50, 56)$	$(59, 65, 75, 85)$	35	$(40, 44, 50, 50)$	$(50, 50, 55, 60)$
16	$(51, 56, 64, 70)$	$(50, 52, 57, 60)$	36	$(70, 76, 85, 85)$	$(30, 34, 43, 47)$
17	$(40, 46, 54, 60)$	$(60, 60, 70, 80)$	37	$(44, 50, 53, 60)$	$(19, 25, 36, 46)$
18	$(55, 60, 65, 70)$	$(50, 54, 61, 67)$	38	$(34, 40, 46, 46)$	$(53, 63, 74, 80)$
19	$(80, 85, 90, 94)$	$(40, 46, 50, 50)$	39	$(40, 45, 51, 60)$	
20	$(80, 84, 90, 90)$	$(44, 50, 56, 66)$	40	$(84, 90, 95, 95)$	

Here we utilize some data given in [25] to compare the opinions of the two experts about the overall impression of the Gamonedo cheese (the trapezoidal fuzzy sets corresponding to their opinions are gathered in Table 2). Thus we have two independent fuzzy samples of sizes $n = 40$ and $m = 38$ comming from the unknown distributions F and G, respectively. Our problem is to check whether there is a general agreement between these two experts. To reach the goal we verify the following null hypothesis $H_0 : F = G$, stating there is no significant difference between experts' opinions, against $H_1 : \neg H_0$ that their opinions on the cheese quality differ.

Substituting data from Table 2 into formula (11) we obtain $t_0 = 1.355$. Then, after combining samples and generating $K = 10\,000$ random permutations we have obtained the p-value of 0.082. Hence, assuming the typical 5% significant level we may conclude that there is no significant difference between the dispersion of experts' opinion on the overal impression of the Gamonedo cheese. In Fig. 3 one can find the empirical null distribution of the permutation test with red vertical line indicating the value t_0 of the test statistic.
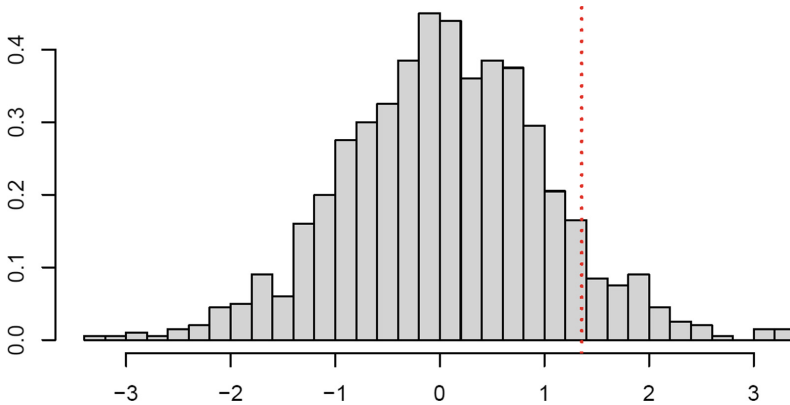
Fig. 3. Empirical null distribution of the permutation test with red vertical line indicating the value of the test statistic.

6   Conclusions

Hypothesis testing with samples which consist of random fuzzy numbers is neither easy nor straightforward. Most of statistical tests developed in this area are based on the bootstrap. In this paper another approach for constructing tests for fuzzy data is proposed. Namely, the two-sample permutation test for dispersion is suggested. Some simulations to illustrate its behavior and to examine its properties are given. Moreover, the case study dedicated to fuzzy rating problem is performed.

The results obtained seem to be promising, but further research including power studies and a comparison with other tests are still intended in the nearest future. In particular, the behavior of the test under strong imbalance in the sample sizes is worth of further examination. Next, we would like to perform an extensive simulation study to compare the performance of our permuatation test and the bootstrap test for the dispersion.

Moreover, some other topics related to the dispersion problem with fuzzy data seem to be of interest. Firstly, we plan to design other two-sample tests for scale, like the permutation test for fuzzy data based on the classical O'Brien test, as well as a permutation test for the homogeneity of more than two fuzzy samples. Secondly, a permutation test for fuzzy data to compare jointly the central tendency and variability of two populations would be of desirable.

References

1. Bickel, P.M., Van Zwet, W.R.: Asymptotic expansion for the power of distribution-free tests in the two-sample problem. Ann. Stat. **6**(987–1004), 1170–1171 (1978)
2. Blanco-Fernández, A., et al.: A distance-based statistic analysis of fuzzy number-valued data (with Rejoinder). Int. J. Approx. Reason. 55, 1487–1501, 1601–1605 (2014)

3. Boos, D.D., Brownie, C.: Comparing variances and other measures of dispersion. Stat. Sci. **19**(4), 571–578 (2004)
4. Diamond, P., Kloeden, P.: Metric spaces of fuzzy sets. Fuzzy Sets Syst. **100**, 63–71 (1999)
5. Gibbons, J.D., Chakraborti, S.: Nonparametric Statistical Inference. Marcel Dekker, Inc., New York (2003)
6. Gil, M.A., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. Metrika **56**, 97–111 (2002)
7. González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.A.: Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. Fuzzy Sets Syst. **157**, 2608–2613 (2006)
8. González-Rodríguez, G., Colubi, A., Gil, M.A.: Fuzzy data treated as functional data. A one-way ANOVA test approach. Comput. Stat. Data Anal. **56**, 943–955 (2012)
9. Good, P.: Permutation, Parametric and Bootstrap Tests of Hypotheses. Springer, Heidelberg (2005). https://doi.org/10.1007/b138696
10. Grzegorzewski, P.: Metrics and orders in space of fuzzy numbers. Fuzzy Sets Syst. **97**, 83–94 (1998)
11. Grzegorzewski, P.: Statical inference about the median from vague data. Control Cybern. **27**, 447–464 (1998)
12. Grzegorzewski, P.: Testing statistical hypotheses with vague data. Fuzzy Sets Syst. **112**, 501–510 (2000)
13. Grzegorzewski, P.: Fuzzy tests - defuzzification and randomization. Fuzzy Sets Syst. **118**, 437–446 (2001)
14. Grzegorzewski, P.: K-sample median test for vague data. Int. J. Intell. Syst. **24**, 529–539 (2009)
15. Grzegorzewski, P.: Two-sample dispersion tests for interval-valued data. In: Medina, J., Ojeda-Aciego, M., Verdegay, J.L., Perfilieva, I., Bouchon-Meunier, B., Yager, R.R. (eds.) IPMU 2018. CCIS, vol. 855, pp. 40–51. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91479-4_4
16. Grzegorzewski, P., Szymanowski, H.: Goodness-of-fit tests for fuzzy data. Inf. Sci. **288**, 374–386 (2014)
17. Lubiano, M.A., Gil, M.A., López-Díaz, M., López-García, M.T.: The λ-mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets Syst. **111**, 307–317 (2000)
18. Lubiano, M.A., Montenegro, M., Sinova, B., de Sáa, S.D.L.R., Gil, M.A.: Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. Eur. J. Oper. Res. **251**, 918–929 (2016)
19. Lubiano, M.A., Salas, A., Carleos, C., de Sáa, S.D.L.R., Gil, M.A.: Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data. Int. J. Approx. Reason. **88**, 128–147 (2017)
20. Marozzi, M.: Levene type tests for the ratio of two scales. J. Stat. Comput. Simul. **81**, 815–826 (2011)
21. Montenegro, M., Colubi, A., Casals, M.R., Gil, M.A.: Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. Metrika **59**, 31–49 (2004)
22. Pan, G.: On a Levene type test for equality of two variances. J. Stat. Comput. Simul. **63**, 59–71 (1999)
23. Pesarin, F.: Multivariate Permutation Tests. Wiley, Hoboken (2001)

24. Puri, M.L., Ralescu, D.A.: Fuzzy random variables. J. Math. Anal. Appl. **114**, 409–422 (1986)
25. Ramos-Guajardo, A.B., Blanco-Fernández, A., González-Rodríguez, G.: Applying statistical methods with imprecise data to quality control in cheese manufacturing. In: Grzegorzewski, P., Kochanski, A., Kacprzyk, J. (eds.) Soft Modeling in Industrial Manufacturing. Studies in Systems, Decision and Control, vol. 183, pp. 127–147. Springer, Heidelberg (2019). https://doi.org/10.1007/978-3-030-03201-2_8
26. Ramos-Guajardo, A.B., Lubiano, M.A.: K-sample tests for equality of variances of random fuzzy sets. Comput. Stat. Data Anal. **56**, 956–966 (2012)
27. Trutschnig, W., González-Rodríguez, G., Colubi, A., Gil, M.A.: A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. Inf. Sci. **179**, 3964–3972 (2009)