

Bootstrap Kuiper Testing of the Identity of 1D Continuous Distributions using Fuzzy Samples

Natalia Nikolova

*Nikola Vaptsarov Naval Academy, V. Drumev 73, Varna 9026, Bulgaria
Australian Maritime College, University of Tasmania, Maritime Way, Newnham TAS 7250, Australia*

Shuhong Chai

Australian Maritime College, University of Tasmania, Maritime Way, Newnham TAS 7250, Australia

Snejana D. Ivanova

Nikola Vaptsarov Naval Academy, V. Drumev 73, Varna 9026, Bulgaria

Krasimir Kolev

Department of Medical Biochemistry, Semmelweis University, 1094 Budapest, Tuzolto u. 37-47, Budapest, Hungary

Kiril Tenekedjiev

*Nikola Vaptsarov Naval Academy, V. Drumev 73, Varna 9026, Bulgaria
Australian Maritime College, University of Tasmania, Maritime Way, Newnham TAS 7250, Australia*

Received 29 July 2015

Accepted 31 October 2015

Abstract

This paper aims to statistically test the null hypothesis H_0 for identity of the probability distribution of one-dimensional (1D) continuous parameters in two different populations, presented by fuzzy samples of i.i.d. observations. A degree of membership to the corresponding population is assigned to any of the observations in the fuzzy sample. The test statistic is the Kuiper's statistic, which measures the identity between the two sample cumulative distribution functions (CDF) of the parameter. A Bootstrap algorithm is developed for simulation-based approximation for the CDF of the Kuiper statistic, provided that H_0 is true. The p_{value} of the statistical test is derived using the constructed conditional distribution of the test statistic. The main idea of the proposed Bootstrap test is that, if H_0 is true, then the two available fuzzy samples can be merged into a unified fuzzy sample. The latter is summarized into a conditional sample distribution of the 1D continuous parameter used for generation of synthetic pairs of fuzzy samples in different pseudo realities. The proposed algorithm has four modifications, which differ by the method to generate the synthetic fuzzy sample and by the type of the conditional sample distribution derived from the unified fuzzy sample used in the generation process. Initial numerical experiments are presented which tend to claim that the four modifications produce similar results.

Keywords: fuzzy samples, percentile Bootstrap procedure, simulation-based algorithm, resemblance of fuzzy samples.

1. Problem Set-Up

Assume that there are two one-dimensional (1D) samples of a continuous parameter that contain respectively n_1 and n_2 observations. Assume also that the observations z_k^1 and z_k^2 respectively of the first and of the second sample belong to two populations, called

Population 1 and Population 2, with degrees of membership respectively μ_k^1 and μ_k^2 . In that sense, the data has fuzzy interpretation, hence we name the samples Fuzzy Sample 1 and Fuzzy Sample 2:

$$Z^1 = \left\{ (z_1^1 - \mu_1^1), (z_2^1 - \mu_2^1), \dots, (z_{n_1}^1 - \mu_{n_1}^1) \right\} \quad (1)$$

$$Z^2 = \left\{ \left(z_1^2 - \mu_1^2 \right), \left(z_2^2 - \mu_2^2 \right), \dots, \left(z_{n_2}^2 - \mu_{n_2}^2 \right) \right\} \quad (2)$$

The objective is to define how the different conditions to form fuzzy samples impact the values of the continuous 1D parameters. This may be interpreted as a case, where the fuzzy samples were formed by two different populations, and it is necessary to test if Population 1 has the same distribution as Population 2.

The most common approximation of the cumulative distribution function (CDF) of a given random variable (r.v.) over a rigid sample is the *empirical distribution function* (ECDF)¹. It only assumes that the observations in the sample are independent and identically distributed (i.i.d.). ECDF approximates the continuous random variable distribution function with a discrete function using the observations in the rigid sample. The assigned probability to each observation is its relative frequency in the sample. Different assumptions for the observations in the sample may result in other approximations of CDF. For example, three different sample approximations with increasing complexity have been reported²: continuous linear, functional continuous linear, and granular continuous linear.

The empirical sample approximation of CDF under the cases of Fuzzy sample Z^1 and Fuzzy sample Z^2 according to equations (1) and (2), may be described as natural generalization of the well-known ECDF formula³:

$$CDF_1(z) = \sum_{\substack{k=1 \\ z_k^1 \leq z}}^{n_1} \mu_k^1 \bigg/ \sum_{k=1}^{n_1} \mu_k^1, \text{ for } z \in (-\infty; +\infty) \quad (3)$$

$$CDF_2(z) = \sum_{\substack{k=1 \\ z_k^2 \leq z}}^{n_2} \mu_k^2 \bigg/ \sum_{k=1}^{n_2} \mu_k^2, \text{ for } z \in (-\infty; +\infty) \quad (4)$$

The same idea can be seen in an implicit form⁴, where the probability mass function (a numerical function indicating the probability that a given discrete r.v. is exactly equal to a given value) has the form corresponding to (3) and (4). Just like the ECDF, the sample approximations (3) and (4) use no assumptions for the type of the approximated CDFs except for the standard requirement that the observations are independent and identically distributed (i.i.d.). Assume that (3) and (4) are called *fuzzy empirical distribution functions* (FECDF).

2. Criteria and Statistical Tests for Identity of 1D Distributions for Two Populations

Most statistical tests that solve the problem of identity of variables distribution in two samples calculate the value s of a given estimator S of the difference in the two sample approximations of CDF and estimate how typical it is assuming that Populations 1 and Population 2 had equal continuous distributions. The estimator S is called *test statistic* and it is a r.v. that tends to increase when the difference between the CDFs of the two samples increases. The null hypothesis H_0 is that the continuous distributions of the two populations are equal, and the alternative hypothesis H_a is that the distributions are different. For the observed value of the test statistic ($S=s$) the probability to reject H_0 that is true (i.e. p_{value}), may be calculated if the conditional distribution of the r.v. S (provided that H_0 is true) is available. Evidently, small values of p_{value} would cast doubt on the validity of H_0 .

Different classes of such measures have been proposed for statistical tests of identity of two continuous distributions. The quadratic class⁵ has its most familiar members to be the Cramér–von Mises statistic and the quadratic Anderson-Darling statistic. The rank class is represented mainly by the Mann-Whitney U statistic and the Wilcoxon T statistic⁶. However, the supremum based class has the dominant position nowadays. The Kolmogorov-Smirnov test statistic KS is probably the most intuitive, straightforward, and commonly used one for the specified problem, being the supremum of the absolute value of the difference between the two approximations of the CDF that result from the samples⁷:

$$KS = \sup_z \left(\left| CDF_1(z) - CDF_2(z) \right| \right) \quad (5)$$

The analytical Kolmogorov-Smirnov test uses (5) as a test statistic and approximates the test p_{value} with a series expansion in the case when (3) and (4) were calculated with unit degrees of membership⁸. An analytical generalization of this solution for the case of fuzzy data sets as in (1) and (2) is not available.

One of the test statistic S with the most desirable properties is the Kuiper statistic (Ku)⁹. It sums the supremum of the positive difference between the two approximations of CDF that result from the samples and the supremum of the negative difference's absolute value between the same:

$$Ku = \sup_z \left(CDF_1(z) - CDF_2(z) \right) + \sup_z \left(CDF_2(z) - CDF_1(z) \right) \quad (6)$$

The Kuiper statistic is a stabilized version of KS , because its sensitivity to the deviation between the distributions is equal for all values of the r.v. This is not the case with the widely used KS statistic, where the deviations are much more difficult to identify at the tails of the interval of z^{10} . It is possible to prove¹¹ that if FECDFs are utilized as approximations in (3) and (4), then the solution of the two optimization tasks over the continuous parameter z in (6) may be replaced with two optimizations over discrete sets with powers respectively n_1 and n_2 :

$$Ku = \max_{k=1,2,\dots,n_1} \left(CDF_1(z_k^1) - CDF_2(z_k^1) \right) + \max_{k=1,2,\dots,n_2} \left(CDF_2(z_k^2) - CDF_1(z_k^2) \right) \quad (7)$$

Furthermore, if (3) and (4) are plugged in (7), then the value of Ku may result directly from the observations in the samples without the need to construct the FECDFs:

$$Ku_r = \max_{i=1,2,\dots,n_1} \left(\frac{\sum_{k=1}^{n_1} \mu_k^1}{\sum_{k=1}^{n_1} \mu_k^1} - \frac{\sum_{k=1}^{n_2} \mu_k^2}{\sum_{k=1}^{n_2} \mu_k^2} \right) + \max_{i=1,2,\dots,n_2} \left(\frac{\sum_{k=1}^{n_2} \mu_k^2}{\sum_{k=1}^{n_2} \mu_k^2} - \frac{\sum_{k=1}^{n_1} \mu_k^1}{\sum_{k=1}^{n_1} \mu_k^1} \right) \quad (8)$$

Formula (8) takes into account the fact that the observations in the samples Z^1 and Z^2 are random, hence Ku is a r.v., and Ku_r is a possible realization. Strictly speaking, the same applies to (6) and (7). The analytical Kuiper test uses (8) for a test statistic and approximates¹² the test p_{value} with a series expansion of scaled Kuiper statistic λ :

$$\lambda = \left(\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} + 0.155 + 0.24 \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} \right) Ku_r \quad (9)$$

$$p_{value} = 2 \sum_{j=1}^{\infty} (4\lambda^2 j^2 - 1) e^{-2\lambda^2 j^2} \quad (10)$$

The resulting p_{value} in (10) is true when the observations in the fuzzy sample (1) and (2) are all with unit degrees of membership. Additionally, the test is asymptotic, which requires a large number of observations in both samples in order to generate reliable results. Finally, analytical generalization of this solution for the case of fuzzy data sets (1) and (2) is not

available. These shortcomings of the analytical Kuiper test also apply to its modification⁹.

Bootstrap simulation is one of the most reliable procedures to calculate the conditional distribution of the statistic S if H_0 is true¹³. The only assumption here is that the observations in the empirical samples are i.i.d. The usefulness of Bootstrap in hypothesis testing with fuzzy data has been discussed in literature¹⁴. The Bootstrap simulation is a computer intensive method since it needs a large number of calculations¹⁵.

There is a Bootstrap procedure to construct the conditional probability distribution of the Kuiper statistic if H_0 is true for the case when each observation in samples (1) and (2) belongs with complete certainty to the respective Population¹⁶. In the current paper these ideas are generalized into a numerical simulation algorithm to calculate p_{value} of the statistical test for identity of the 1D continuous distributions of the two populations represented by fuzzy empirical samples as in (1) and (2).

3. Generating Synthetic Fuzzy Samples in a Bootstrap Simulation

Assume that the continuous distribution of the two populations are the same and equal to $CDF_{tr}(\cdot)$ (which would be the true CDF, if H_0 is true). Then the true value of the Kuiper statistic shall be $Ku_{tr}=0$. These two equal distributions are not available to the analyst, but are realized statistically by the pair of samples $(Z^1 - Z^2)$, given in (1) and (2). Using (8) over these samples, it is possible to calculate the real estimate Ku_r of the Kuiper statistic. $CDF_{tr}(\cdot)$ allows to generate countless number of *hypothetic sample pairs*, similar to the sample pair $(Z^1 - Z^2)$: $(Z_1^1 - Z_1^2)$, $(Z_2^1 - Z_2^2)$, $(Z_3^1 - Z_3^2)$, The analyst could calculate the hypothetical realizations Ku_1 , Ku_2 , Ku_3 , ... of the Kuiper statistic based on each of the hypothetical sample pairs. The hypothetical realizations Ku_1 , Ku_2 , Ku_3 , ... would reveal the probability distribution of the sample numerical characteristics Ku , and Ku_r would be a possible realization. The difference $root = Ku - Ku_{tr} = Ku - 0 = Ku$ is an r.v., called *pivot*¹³ or *root*¹² and its distribution entirely describes the quantitative uncertainty in the real parameters Ku_r . The main task of simulation techniques is to describe the distribution of *root* if $CDF_{tr}(\cdot)$ and the hypothetical sample pairs $(Z_1^1 - Z_1^2)$, $(Z_2^1 - Z_2^2)$, $(Z_3^1 - Z_3^2)$, ... are not known. Since the true $CDF_{tr}(\cdot)$ is unknown, then it may be replaced by its estimate¹⁷. If H_0 is true, then $CDF_{tr}(\cdot)$ is statistically realized $(n_1 + n_2)$ times in the samples (1) and (2). Let's unite the samples Z^1 and Z^2 into a single sample that contains $(n_1 + n_2)$ observations:

$$Z^{1-2} = \left\{ \left(z_1^1 - \mu_1^1 \right), \left(z_2^1 - \mu_2^1 \right), \dots, \left(z_{n_1}^1 - \mu_{n_1}^1 \right), \right. \\ \left. \left(z_1^2 - \mu_1^2 \right), \left(z_2^2 - \mu_2^2 \right), \dots, \left(z_{n_2}^2 - \mu_{n_2}^2 \right) \right\} = \\ \left\{ \left(z_1^{1-2} - \mu_1^{1-2} \right), \left(z_2^{1-2} - \mu_2^{1-2} \right), \dots, \right. \\ \left. \left(z_{n_1+n_2}^{1-2} - \mu_{n_1+n_2}^{1-2} \right) \right\} \quad (11)$$

Then $CDF_{tr}(\cdot)$ may be intuitively assessed with two different sample step function approximations.

The first sample approximation is ECDF based on (11):

$$CDF_{1-2}(z) = \frac{1}{n_1 + n_2} \sum_{\substack{k=1 \\ z_k^{1-2} \leq z}}^{n_1+n_2} 1, \text{ for } z \in (-\infty; +\infty) \quad (12)$$

This approximation is employed by the classical Bootstrap approach, where each observation from the unified sample (11) has equal chance of being generated, therefore the Bootstrap generation is performed using sampling with replacement from (11).

The second sample approximation is FECDF based on (11):

$$CDF_{1-2}(z) = \sum_{\substack{k=1 \\ z_k^{1-2} \leq z}}^{n_1+n_2} \mu_k^{1-2} / \sum_{k=1}^{n_1+n_2} \mu_k^{1-2}, \text{ for } z \in (-\infty; +\infty) \quad (13)$$

Here, the Bootstrap approach is modified so that each observation from the unified sample (11) has a probability to be generated, proportional to its degree of membership $P(Z = z_k^{1-2}) = \mu_k^{1-2} / \sum_{k=1}^{n_1+n_2} \mu_k^{1-2}$, for $k=1, 2, \dots, (n_1+n_2)$. The approximation (13) is the cumulative distribution function of a discrete r.v. with possible values coinciding with the different observations in the fuzzy sample (11). The assigned probability to each value is proportional to the sum of the degrees of membership of the instances of that value in the fuzzy sample (11).

Assume that in a computer-simulated environment (called 'Bootstrap world'¹⁸), the true $CDF_{tr}(\cdot)$ of the two populations is replaced by one of the two sample approximations $CDF_{1-2}(\cdot)$ from (12) or (13). Since in the Bootstrap world the two populations have equal distributions, then the value of the Kuiper statistic would be $Ku_{cw}=0$. In the Bootstrap world, the known $CDF_{1-2}(\cdot)$ allows to generate M synthetic pairs of fuzzy samples $(Z_q^{1,s} - Z_q^{2,s})$, for $q=1, 2, \dots, M$. The synthetic

pairs of fuzzy samples (again) can be generated using two intuitive methods ensuring that they resemble the original pair $(Z^1 - Z^2)$.

A) *Equal-size generation*: The method generates synthetic pairs of fuzzy samples with the same sizes as the original fuzzy samples (1) and (2):

$$Z_q^{1,s} = \left\{ \left(z_{1,q}^{1,s} - \mu_{1,q}^{1,s} \right), \left(z_{2,q}^{1,s} - \mu_{2,q}^{1,s} \right), \dots, \left(z_{n_1,q}^{1,s} - \mu_{n_1,q}^{1,s} \right) \right\} \quad (14)$$

$$Z_q^{2,s} = \left\{ \left(z_{1,q}^{2,s} - \mu_{1,q}^{2,s} \right), \left(z_{2,q}^{2,s} - \mu_{2,q}^{2,s} \right), \dots, \left(z_{n_2,q}^{2,s} - \mu_{n_2,q}^{2,s} \right) \right\} \quad (15)$$

The equal-size generation is used in the classical Bootstrap procedure, but also in the entire Monte Carlo simulation, when the true distributions are considered known.

B) *Quasi-equal-information generation*: If, for example, an observation in a fuzzy sample has a degree of membership to the Population equal to 1, then the information it contains should be equal to the information supplied by two observations of the same fuzzy sample with degrees of membership to the Population equal to 0.65 and 0.35. It follows that the information content in a fuzzy sample can be measured with the sum of the degrees of membership of its observations. The proposed method generates synthetic pairs of fuzzy samples with approximately the same information content as the original fuzzy samples (1) and (2). This method is realized in three steps.

B1) in the first step, two initial synthetic fuzzy samples with preselected numbers of observations have to be generated, using the known $CDF_{1-2}(\cdot)$:

$$Z_{q,ini}^{1,s} = \left\{ \left(z_{1,q}^{1,s} - \mu_{1,q}^{1,s} \right), \left(z_{2,q}^{1,s} - \mu_{2,q}^{1,s} \right), \dots, \left(z_{g_1,q}^{1,s} - \mu_{g_1,q}^{1,s} \right) \right\} \quad (16)$$

$$Z_{q,ini}^{2,s} = \left\{ \left(z_{1,q}^{2,s} - \mu_{1,q}^{2,s} \right), \left(z_{2,q}^{2,s} - \mu_{2,q}^{2,s} \right), \dots, \left(z_{g_2,q}^{2,s} - \mu_{g_2,q}^{2,s} \right) \right\} \quad (17)$$

The counts g_1 and g_2 are determined as the minimal sizes, which ensure that the initial synthetic fuzzy samples would contain not less information than the original fuzzy samples (1) and (2):

$$g_1 = \left\lceil \sum_{k=1}^{n_1} \mu_k^1 / \min \{ \mu_1^{1-2}, \mu_2^{1-2}, \dots, \mu_{n_1+n_2}^{1-2} \} \right\rceil \quad (18)$$

$$g_2 = \left\lceil \sum_{k=1}^{n_2} \mu_k^2 / \min \{ \mu_1^{1-2}, \mu_2^{1-2}, \dots, \mu_{n_1+n_2}^{1-2} \} \right\rceil \quad (19)$$

Here $\lceil x \rceil$ stands for the minimum integer, which is not less than x . In fact $\lceil \cdot \rceil$ rounds the values upwards.

B2) in the second step, the sizes $n_{1,q}$ and $n_{2,q}$ of the final synthetic fuzzy samples have to be evaluated in such a

way that the first $n_{1,q}$ and $n_{2,q}$ observations of (16) and (18) have information content as close as possible to the information content of the original fuzzy samples (1) and (2):

$$n_{1,q} = \arg \min_{i=1,2,\dots,g_1} \left\{ \sum_{k=1}^i \mu_{k,q}^{1,s} - \sum_{k=1}^{n_1} \mu_k^1 \right\} \quad (20)$$

$$n_{2,q} = \arg \min_{i=1,2,\dots,g_2} \left\{ \sum_{k=1}^i \mu_{k,q}^{2,s} - \sum_{k=1}^{n_2} \mu_k^2 \right\} \quad (21)$$

B3) in the third step, the final synthetic fuzzy samples are obtained by cutting down the two generated initial synthetic fuzzy samples to their first $n_{1,q}$ and $n_{2,q}$ fuzzy observations:

$$Z_q^{1,s} = \left\{ \left(z_{1,q}^{1,s} - \mu_{1,q}^{1,s} \right), \left(z_{2,q}^{1,s} - \mu_{2,q}^{1,s} \right), \dots, \left(z_{n_{1,q},q}^{1,s} - \mu_{n_{1,q},q}^{1,s} \right) \right\} \quad (22)$$

$$Z_q^{2,s} = \left\{ \left(z_{1,q}^{2,s} - \mu_{1,q}^{2,s} \right), \left(z_{2,q}^{2,s} - \mu_{2,q}^{2,s} \right), \dots, \left(z_{n_{2,q},q}^{2,s} - \mu_{n_{2,q},q}^{2,s} \right) \right\} \quad (23)$$

4. Assessing p_{value} in a Bootstrap Kuiper test with Fuzzy Samples

Regardless of whether the synthetic samples were derived as in (14) and (15) using equal-size generation or as in (22) and (23) using quasi-equal-information generation, for the q -th pair of fuzzy samples it is possible to calculate the synthetic Kuiper statistic Ku_q^s using (8), which is a realization of the sample numerical characteristic in the Bootstrap world Ku^s :

$$Ku_q^s = \max_{i=1,2,\dots,n_{1,g}} \left(\frac{\sum_{k=1}^{n_{1,g}} \mu_{k,q}^{1,s}}{\sum_{k=1}^{n_{1,g}} \mu_{k,q}^{1,s}} - \frac{\sum_{k=1}^{n_{2,g}} \mu_{k,q}^{2,s}}{\sum_{k=1}^{n_{2,g}} \mu_{k,q}^{2,s}} \right) + \max_{i=1,2,\dots,n_{2,g}} \left(\frac{\sum_{k=1}^{n_{2,g}} \mu_{k,q}^{2,s}}{\sum_{k=1}^{n_{2,g}} \mu_{k,q}^{2,s}} - \frac{\sum_{k=1}^{n_{1,g}} \mu_{k,q}^{1,s}}{\sum_{k=1}^{n_{1,g}} \mu_{k,q}^{1,s}} \right) \quad (24)$$

In (24) $n_{1,g}$ and $n_{2,g}$ are the generalized sizes of the synthetic fuzzy samples in the q^{th} synthetic pair:

$$n_{1,g} = \begin{cases} n_1 & \text{for 'equal-size generation'} \\ n_{1,q} & \text{for 'quasi-equal-information generation'} \end{cases} \quad (25)$$

$$n_{2,g} = \begin{cases} n_2 & \text{for 'equal-size generation'} \\ n_{2,q} & \text{for 'quasi-equal-information generation'} \end{cases} \quad (26)$$

After M computer-simulated realities (called pseudo realities¹²), a synthetic sample S_{Ku} containing M number of instances of the r.v. Ku^s is generated:

$$S_{Ku} = \{Ku_1^s, Ku_2^s, \dots, Ku_M^s\} \quad (27)$$

If M is a great number (for example 100000), then the cumulative distribution of the r.v. Ku^s may be constructed with high precision as ECDF over (27):

$$CDF_{Ku}(k_u) = \frac{1}{M} \sum_{\substack{q=1 \\ Ku_q^s \leq k_u}}^M 1, \text{ for } k_u \in (-\infty; +\infty) \quad (28)$$

The difference $Ku^s - Ku_{cw} = Ku^s - 0 = Ku^s$ is a r.v. called $root_{bw}$. The main assumption of the Bootstrap technique is that the distribution of $root_{bw}$ in the Bootstrap world is the same as the distribution of $root$ in the real world¹⁵.

Since in that case, $root$ coincides with Ku , and $root_{bw}$ coincides with Ku^s , then the necessary distribution of the sample numerical characteristic Ku coincides with the constructed distribution of Ku^s in (28). So the p_{value} of the test is the relative number of synthetic Kuiper statistic higher than the real estimate Ku_r derived in (8):

$$p_{value} = 1 - CDF_{Ku}(Ku_r) = \frac{1}{M} \sum_{\substack{q=1 \\ Ku_q^s > Ku_r}}^M 1 \quad (29)$$

The necessary p_{value} of the Bootstrap Kuiper test over fuzzy samples may be calculated as follows.

Bootstrap algorithm for estimation of p_{value} for the Kuiper test over fuzzy samples

1. Select the sample approximation of CDF using the unified sample:

$$Fdist = \begin{cases} 0 & \text{for ECDF} \\ 1 & \text{for FECDF} \end{cases}$$

2. Select the generation method for synthetic pairs of fuzzy samples:

$$Fgen = \begin{cases} 0 & \text{for 'equal-size generation'} \\ 1 & \text{for 'quasi-equal-information generation'} \end{cases}$$

3. Select the count of pseudo realities M ;
4. Estimate the actual realization of the Kuiper statistic Ku_r according to (8);
5. Form the unified fuzzy sample Z^{1-2} according to (11);
6. If $Fgen$ is 1, then determine the fixed sizes of the initial synthetic fuzzy samples:
 - 6.1. Calculate the size g_1 of the first initial synthetic samples $Z_{q,ini}^{1,s}$ according to (18);
 - 6.2. Calculate the size g_2 of the second initial synthetic samples $Z_{q,ini}^{2,s}$ according to (19);
7. Initialize pseudo reality cycle: $q=1$;
8. Form the synthetic fuzzy sample $Z_q^{1,s}$:
 - 8.1. If $Fgen$ is 1, then go to step 8.8.;
 - 8.2. Initialize synthetic observation cycle: $k=1$;
 - 8.3. If $Fdist$ is 0, then draw the k^{th} observation of $Z_q^{1,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (12):
 - 8.3.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;
 - 8.3.2. find the number in Z^{1-2} of the k^{th} observation of $Z_q^{1,s}$: $numt = \lceil rand \times (n_1 + n_2) \rceil$;
 - 8.3.3. remember the k^{th} observation of $Z_q^{1,s}$: $(z_{k,q}^{1,s} - \mu_{k,q}^{1,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2})$;
 - 8.4. If $Fdist$ is 1, then draw the k^{th} observation of $Z_q^{1,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (13):
 - 8.4.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;
 - 8.4.2. find the number in Z^{1-2} of the k^{th} observation of $Z_q^{1,s}$: $numt$ is the minimal i from the set $\{1, 2, \dots, n_1 + n_2\}$ which satisfies the condition $\left(\sum_{k=1}^i \mu_k^{1-2} \right) > rand$;
 - 8.4.3. remember the k^{th} observation of $Z_q^{1,s}$: $(z_{k,q}^{1,s} - \mu_{k,q}^{1,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2})$;
 - 8.5. Move to the next observation: $k=k+1$;
 - 8.6. Check whether there are enough generated observations for $Z_q^{1,s}$: if $k \leq n_1$, then go to step 8.3;
 - 8.7. Form the synthetic fuzzy sample $Z_q^{1,s}$ according to (14) and go to step 9;

- 8.8. Start initial synthetic observation cycle: $k=1$;
- 8.9. If $Fdist$ is 0, then draw the k^{th} observation of $Z_{q,ini}^{1,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (12):
 - 8.9.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;
 - 8.9.2. find the number in Z^{1-2} of the k^{th} observation of $Z_{q,ini}^{1,s}$: $numt = \lceil rand \times (n_1 + n_2) \rceil$;
 - 8.9.3. remember the k^{th} observation of $Z_{q,ini}^{1,s}$: $(z_{k,q}^{1,s} - \mu_{k,q}^{1,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2})$;
- 8.10. If $Fdist$ is 1, then draw the k^{th} observation of $Z_{q,ini}^{1,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (13):
 - 8.10.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;
 - 8.10.2. find the number in Z^{1-2} of the k^{th} observation of $Z_{q,ini}^{1,s}$: $numt$ is the minimal i from the set $\{1, 2, \dots, n_1 + n_2\}$ which satisfies the condition $\left(\sum_{k=1}^i \mu_k^{1-2} \right) > rand$;
 - 8.10.3. remember the k^{th} observation of $Z_{q,ini}^{1,s}$: $(z_{k,q}^{1,s} - \mu_{k,q}^{1,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2})$;
- 8.11. Move to the next observation: $k=k+1$;
- 8.12. Check whether there are enough generated observations for $Z_{q,ini}^{1,s}$: if $k \leq g_1$, then go to step 8.9;
- 8.13. Form the initial synthetic fuzzy sample $Z_{q,ini}^{1,s}$ according to (16);
- 8.14. Calculate the size $n_{1,q}$ of the final synthetic fuzzy sample $Z_q^{1,s}$ according to (20);
- 8.15. Form the final synthetic fuzzy sample $Z_q^{1,s}$ according to (22) by cutting down $Z_{q,ini}^{1,s}$ to its first $n_{1,q}$ fuzzy observations;
9. Form the synthetic fuzzy sample $Z_q^{2,s}$:
 - 9.1. If $Fgen$ is 1, then go to step 9.8;
 - 9.2. Initialize synthetic observation cycle: $k=1$;
 - 9.3. If $Fdist$ is 0, then draw the k^{th} observation of $Z_q^{2,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (12):
 - 9.3.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;
 - 9.3.2. find the number in Z^{1-2} of the k^{th} observation of $Z_q^{2,s}$: $numt = \lceil rand \times (n_1 + n_2) \rceil$;

9.3.3. remember the k^{th} observation of $Z_q^{2,s}$:

$$(z_{k,q}^{2,s} - \mu_{k,q}^{2,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2});$$

9.4. If $Fdist$ is 1, then draw the k^{th} observation of $Z_q^{2,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (13):

9.4.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;

9.4.2. find the number in Z^{1-2} of the k^{th} observation of $Z_q^{2,s}$: $numt$ is the minimal i from the set $\{1, 2, \dots, n_1 + n_2\}$

which satisfies the condition $\left(\sum_{k=1}^i \mu_k^{1-2}\right) > rand$;

9.4.3. remember the k^{th} observation of $Z_q^{2,s}$:

$$(z_{k,q}^{2,s} - \mu_{k,q}^{2,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2});$$

9.5. Move to the next observation: $k=k+1$;

9.6. Check whether there are enough generated observations for $Z_q^{2,s}$: if $k \leq n_2$, then go to step 9.3;

9.7. Form the synthetic fuzzy sample $Z_q^{2,s}$ according to (15) and go to step 10;

9.8. Start initial synthetic observation cycle: $k=1$;

9.9. If $Fdist$ is 0, then draw the k^{th} observation of $Z_{q,ini}^{2,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (12):

9.9.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;

9.9.2. find the number in Z^{1-2} of the k^{th} observation of $Z_{q,ini}^{2,s}$: $numt = \lceil rand \times (n_1 + n_2) \rceil$;

9.9.3. remember the k^{th} observation of $Z_{q,ini}^{2,s}$:

$$(z_{k,q}^{2,s} - \mu_{k,q}^{2,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2});$$

9.10. If $Fdist$ is 1, then draw the k^{th} observation of $Z_{q,ini}^{2,s}$ from the unified fuzzy sample Z^{1-2} by generation according to (13):

9.10.1. generate a pseudo-random number $rand \in [0;1)$ from an even distribution;

9.10.2. find the number in Z^{1-2} of the k^{th} observation of $Z_{q,ini}^{2,s}$: $numt$ is the minimal i from the set $\{1, 2, \dots, n_1 + n_2\}$ which satisfies the condition $\left(\sum_{k=1}^i \mu_k^{1-2}\right) > rand$;

9.10.3. remember the k^{th} observation of $Z_{q,ini}^{2,s}$:

$$(z_{k,q}^{2,s} - \mu_{k,q}^{2,s}) = (z_{numt}^{1-2} - \mu_{numt}^{1-2});$$

9.11. Move to the next observation: $k=k+1$;

9.12. Check whether there are enough generated observations for $Z_{q,ini}^{2,s}$: if $k \leq g_2$, then go to step 9.9;

9.13. Form the initial synthetic fuzzy sample $Z_{q,ini}^{2,s}$ according to (17);

9.14. Calculate the size $n_{1,q}$ of the final synthetic fuzzy sample $Z_q^{2,s}$ according to (21);

9.15. Form the final synthetic fuzzy sample $Z_q^{2,s}$ according to (23) by cutting down $Z_q^{2,s}$ to its first $n_{2,q}$ fuzzy observations;

10. Calculate the generalized number $n_{1,g}$ of the fuzzy observations in $Z_q^{1,s}$ according to (25);

11. Calculate the generalized number $n_{2,g}$ of the fuzzy observations in $Z_q^{2,s}$ according to (26);

12. Calculate the Kuiper's statistic Ku_q^s in the q^{th} pseudo reality according to (24);

13. Move to the next pseudo reality: $q=q+1$;

14. Check for end of simulation: if $q \leq M$, go to step 8;

15. Form the synthetic sample S_{Ku} containing M instances of the r.v. Ku^s according to (27);

16. Calculate p_{value} according to (29).

To clarify the procedure, the steps in the above algorithm are organized in four parallel schemes, one for each of the Kuiper test modifications according to the values $Fdist$ and $Fgen$ in the presented unified algorithm (see Fig. 1 to Fig. 4).

5. Interpretation of the Set-up Problem

The number of studies, where fuzzy techniques have been applied, in the statistical analysis is continuously growing under the Soft Computing paradigm. The objectives of such studies involve four different purposes, the first and the last being¹⁹: "... (i) to introduce new data analysis problems in which the objective involves either fuzzy relationships or fuzzy terms; ... (iv) to incorporate fuzzy sets to help in solving traditional statistical problems with non-fuzzy data". The presented research represents the outlined trend. Depending on the set-up interpretation, this work can contribute in reaching both of the above-mentioned purposes.

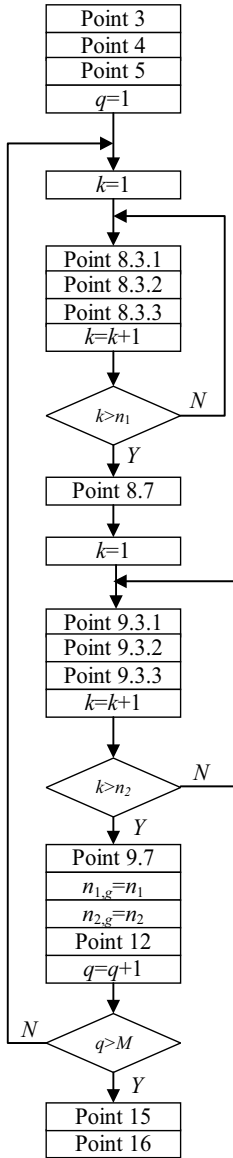


Fig.1. Scheme for p_{value} Bootstrap calculation with 'equal-size' generation using ECDF

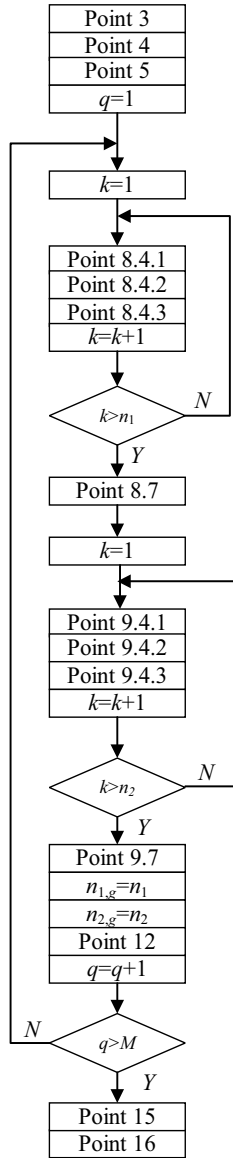


Fig.2. Scheme for p_{value} Bootstrap calculation with 'equal-size' generation using FECDF

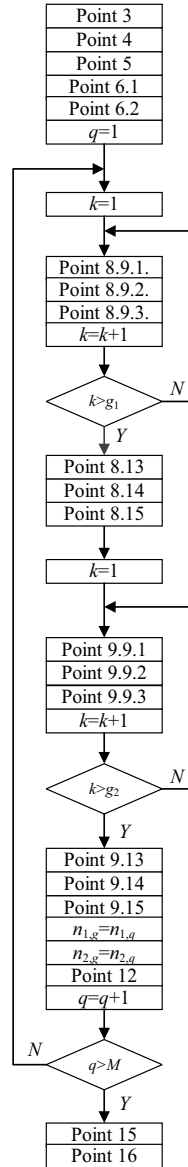


Fig.3. Scheme for p_{value} Bootstrap calculation with 'quasi-equal information' generation using ECDF

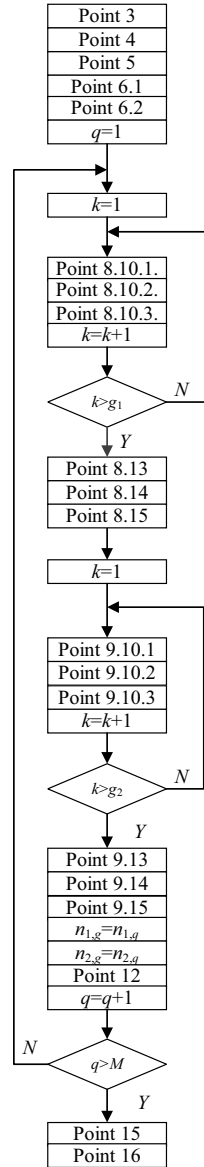


Fig.4. Scheme for p_{value} Bootstrap calculation with 'quasi-equal information' generation using FECDF

5.1. Solving Data Analysis Problems Involving Fuzzy Relations

There are generally two wide spread interpretations of fuzzy data measured in random experiments: epistemic and physical^{20,21}. Under the epistemic interpretation, it is considered that the result of each experiment is a crisp original r.v. value, which, however, is partially observed

and cannot be precisely measured. Almost every classical statistic can be fuzzified using the extension principle²². As stated in literature²³: "in general, fuzziness and errors are superimposed". Under the physical interpretation, it is assumed that data is intrinsically fuzzy and therefore the result of any experiment is a fuzzy value. When discussing fuzzy

random variables, it is claimed²⁴ that the latter were invented for the case when random mechanisms are generating imprecisely-valued fuzzy data. The two interpretations are not competing, but rather complementary and choosing which to adopt in each particular case depends more on the problem itself, rather than on the attitude of the researcher.

The proposed Bootstrap Kuiper test that uses fuzzy samples is most useful under the physical interpretation. For example, consider patients with ischemic heart disease also experiencing mild to moderate ischemic mitral regurgitation. These patients are divided in two groups²⁵ – A and B. Group A consists of patients who were subjected to combined myocardial revascularization with mitral valve repair, while group B consists of patients who were subjected to isolated myocardial revascularization. One of the important continuous parameters, indicating medical status, is the ejection fraction (EF), which shows what percentage of maximal blood volume in the heart is ejected. Since both groups are quite heterogeneous, each is further divided into two comparatively homogeneous subgroups – A1, A2, B1 and B2. Subgroups A1 and B1 include patients with comparatively preserved medical condition, while subgroups A2 and B2 contain patients with worse medical condition. Obviously, the medical condition of a patient from group A is intrinsically fuzzy, because it can not only take two states (A1 or A2), but the whole spectrum between A1 and A2. The same applies to group B. Each patient is assigned a degree of membership to its own subgroup either by the cardiac team or by a fuzzy classifier²⁶. If the problem is to test the identity of EF distributions in subgroups A1 and B1 late after surgical operation, then two fuzzy samples can be formed. The first fuzzy sample contains the measured EF value for each patient in group A1 along with the assigned degree of membership of the patient to A1. The second fuzzy sample has the same content, but refers to B1. Then, the set-up is the same as the one in Section 1.

5.2. Incorporating Fuzzy Sets in Statistical Problems with Non-Fuzzy Data

The set-up in Section 1 and the proposed solution can be applied to problems with non-fuzzy data by using the form of fuzzy samples (1) and (2). For example, in order to improve the understanding of the pathomechanism of ischemic arterial (coronary or peripheral) disease, thrombus samples of patients are collected and processed with two techniques (electron microscopic and immunohistochemical) to acquire quantitative data on thrombus composition. For each of the techniques, between 2 and 5 different sections of each thrombus are

used and in each section between 2 and 5 regions of interest are evaluated. The electron microscopic measurements generate data on fibrin fiber diameter and thrombus occupancy by fibrin and blood cells (red blood cells, leukocytes and platelets). The immunohistochemical measurements generate data on the ratio of fibrin and platelet antigens²⁷. The data is analysed using hypothesis testing for differences between groups of patients with common clinical features in order to draw general conclusions on the characteristics of the disease. However, a single thrombus from a single patient is exposed to effects that vary in space (e.g. blood flow in different regions of the thrombus). This results in a high degree of biological diversity reflected in the heterogeneity of the data measured in different areas of a thrombus. A statistical analytical procedure of hypothesis testing could account for this heterogeneity treating all data as separate observations characterizing the disease, if intra-individual heterogeneity is neglected. However, because of different size of the thrombus samples and consequent difference in the number of available measured data from each patient, this approach distorts the role of inter-individual differences ascribing higher weight for patients with larger sets of measured data. If the inter-individual differences are to be considered as a factor in the disease mechanism, all intra-individual data should be taken into account in the statistical analysis as a single observation originating as a dataset composed of discrete subsets of actually measured data with appropriate weighing factor reflecting the total number of measurements taken from each thrombus. For example, red blood cells are known to modify the fibrin structure causing formation of thinner fibers²⁸. The red blood cell occupancy in different sections of the same thrombus can vary between 7% and 21%²⁷. If these are considered as separate observations, this artificially increases the number of clots with thin or with thick fibers taken into consideration. That is why, a good way forward would be to form a sample with red blood cell occupancy in all coronary thrombus sections available, but to weight each measured data for the patient number j with $1/t_j$, where t_j is the count of the thrombus sections for the patient number j . Similar sample could be formed for the peripheral thrombus sections. By doing so the samples would reflect the true inter-individual heterogeneity in red blood cell occupancy, because each patient would contribute equally to the result. There is nothing fuzzy in the observations or in the weights, but it is convenient to use the set-up from Section 1 assigning $1/t_j$ to the membership degree of the corresponding observation (red blood cell occupancy in a thrombus sections) in the 'fuzzy' samples.

6. Numerical Examples

Numerical example 1: Assume there are two fuzzy samples $Z^1 = \{(15.25-1), (17.16-1), (11.55-1), (17.28-1), (23.88-1), (20.11-1)\}$ and $Z^2 = \{(15.19-1), (12.28-0.5), (31.38-0.7), (18.80-1), (15.30-0.6), (19.02-0.8), (18.97-1), (21.28-1)\}$. The rigid samples containing the same observations are presented in earlier works¹⁶. The test statistic Ku_r is 0.4242 according to (9). Each of the four modifications of the proposed algorithm are realized with $M=1000$ pseudo realities. The corresponding Bootstrap generated CDFs of the Kuiper's statistic are shown on Fig.5. The p_{value} for each of the Kuiper test modifications is: a) 0.649 for Bootstrap with quasi-equal-information generation using ECDF; b) 0.662 for Bootstrap with quasi-equal-information generation using FECDF; c) 0.666 for Bootstrap with equal-size generation using ECDF; d) 0.625 for Bootstrap with equal-size generation using FECDF. Obviously, the data in the fuzzy samples does not contradict the null hypothesis for any reasonable significance level.

In fact, the observations in the two samples are generated from equal continuous distributions of the two populations. The unified distribution is a mixture of 40% normal distribution with a mathematical expectation 14 and with variance 16, and 60% χ^2 with

two degrees of freedom, with expectation 19 and variance 16. Since here, $CDF_v(z)$ is known, then the true distribution of the Kuiper statistic is derived using Monte Carlo simulation with 1000 hypothetic samples. In the q^{th} pseudo reality, there are respectively 6 and 8 observations generated from $CDF_v(z)$. Each of the 14 observations has an associated degree of membership to the respective population from the given Z^1 and Z^2 . In that way, a synthetic pair in the analyzed pseudo reality is formed from fuzzy samples $(Z_q^{1,MK} - Z_q^{2,MK})$, for $q=1, 2, \dots, 1000$. Using the q^{th} pair of samples it is possible to calculate the realization Ku_q^{MC} of the Kuiper statistic in a pseudo reality. Based on the resulting pseudo real realizations $Ku_1^{MC}, Ku_2^{MC}, \dots, Ku_{1000}^{MC}$ it is possible to construct the CDF of the Kuiper statistic using Monte Carlo simulation. The latter is presented on each of the four graphics on Fig. 5. It is obvious that regardless of the small number of fuzzy samples, the approximation of the distribution of the Kuiper statistic is comparatively close to the real Monte Carlo approximation. In the same time, there is no substantial difference between the results from the four versions of the proposed algorithm.

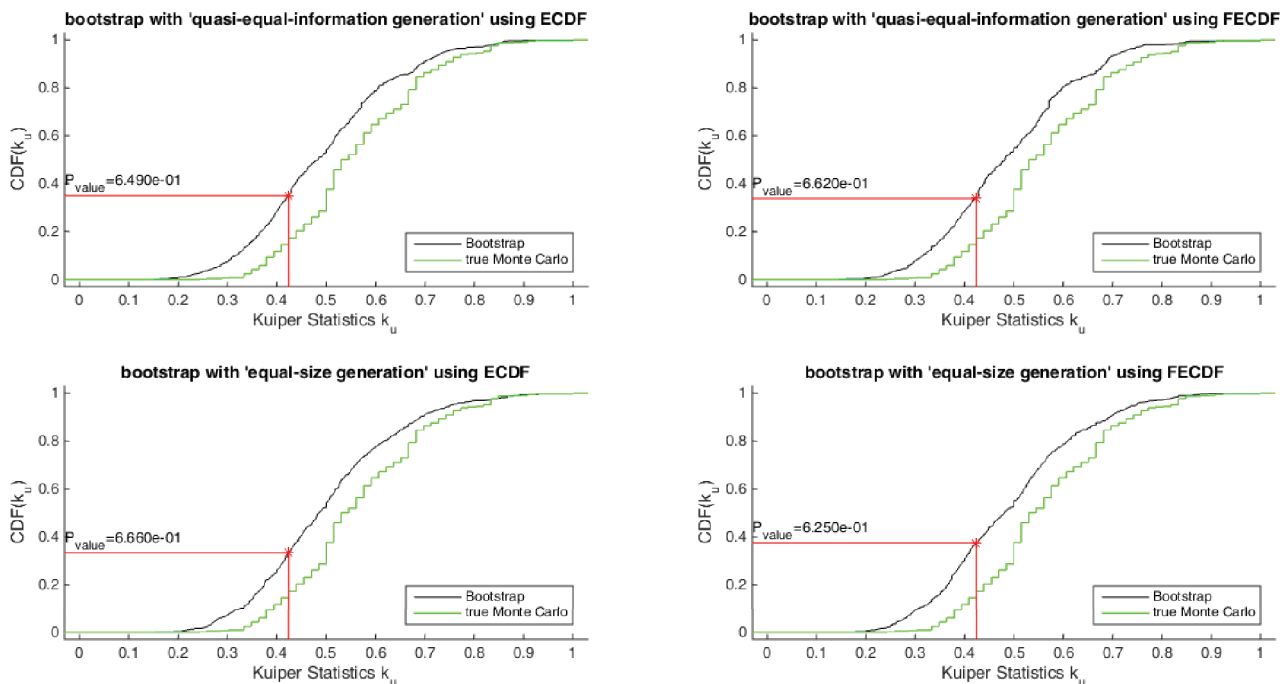


Fig. 5. Bootstrap generated CDFs of the Kuiper's statistic for the fuzzy samples in Numerical example 1, along with the true Monte-Carlo empirical CDF of the same data

Numerical example 2: Let $Z^1 = \{(12.38-1), (13.01-1), (11.14-1), (13.05-1), (15.25-1), (14.00-1)\}$ and Z^2 be as in Numerical Example 1. The rigid samples containing the same observations are presented in earlier works¹⁶, where it is stated that in the rigid sample case, the analytical Kuiper test fails to prove the difference between the population CDFs. The test statistic calculated using (9) is $Ku_r = 0.7727$. Each of the four modifications of the proposed algorithm is realized with

$M=1000$ pseudo realities and the corresponding Bootstrap generated CDFs of the Kuiper statistic are shown on Fig. 6. The p_{value} for each of the Kuiper test modifications is: a) 0.022 for Bootstrap with quasi-equal-information generation using ECDF; b) 0.026 for Bootstrap with quasi-equal-information generation using FECDF; c) 0.049 for Bootstrap with equal-size generation using ECDF; d) 0.039 for Bootstrap with equal-size generation using FECDF.

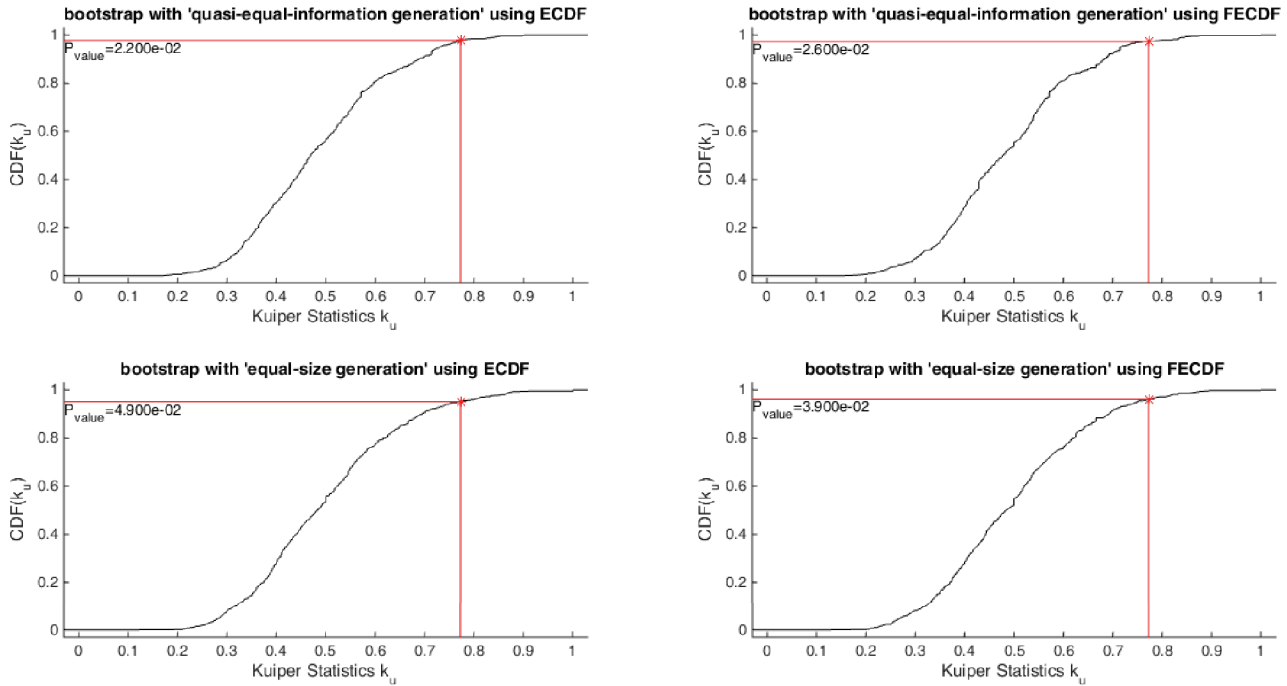


Fig. 6. Bootstrap generated CDFs of the Kuiper's statistic for the fuzzy samples in Numerical example 2

At 5% significance level, the null hypothesis has to be rejected by each of the four modifications of the Bootstrap Kuiper test with fuzzy samples. There is no recognizable difference between the results of the four modifications of the proposed algorithm, in the same way as in Numerical example 1.

7. Conclusion

It is reasonable to ask which of the four proposed modifications is to be preferred, but that question has no definite answer, yet. Some discussion without empirical background may be offered here.

First of all, we can compare the generation methods. The 'equal-size' generation method is simple. It is traditionally well established and theoretically sound at

least for the special case of crisp samples, so our guess is that the end users would trust it more. Unfortunately, there are no generalizations of the basic Bootstrap theorems for the case with fuzzy samples (1) and (2). In the same time, the 'quasi-equal-information' method may be expected to produce more stable generation without the expected odd interruptions and outliers which the 'equal-size' generation method is prone to give, due to the possibility to produce vast difference in the information content in different pseudo realities.

Second of all, we can discuss the distribution used for generation. The CDF of any population can be better approximated by the FECDF of a fuzzy sample, compared to approximation by the ECDF of the same sample. That is why the proposed algorithm always calculates the Kuiper statistic as special difference of

the FECDFs of two fuzzy samples, those being either the original ones or the synthetic samples. At first glance it appears that the generation of the synthetic samples in different pseudo realities should utilize the FECDF, rather than the ECDF of the unified fuzzy sample. The last statement is, however, flawed. A low degree of membership of a specific fuzzy observation, included in a synthetic fuzzy sample, ensures less influence on the sample FECDF of the parameter in the pseudo reality. Using the FECDF of the unified fuzzy sample for generation of the synthetic samples ensures that this synthetic observation has even less chance to be included in the synthetic fuzzy sample. So, the low degree of membership is in a way double-accounted. That is why, it can be expected that the modifications, which generate the synthetic fuzzy samples from the ECDF of the unified fuzzy sample, would produce better results than the corresponding ones, which utilize the FECDF of the unified fuzzy sample for generation. Similar statement can be discovered in literature²⁹. However, the two numerical examples, presented here, do not support this claim, and produce similar results.

From the above considerations, it is obvious that new extensive numerical simulations and new theoretical advances are needed to clarify which of the four modifications should be trusted. It is quite possible to have problem- and/or interpretation-dependent answer. Having said that, our recommended selection at this stage would be Bootstrap Kuiper test with 'quasi-equal information' generation using ECDF.

Another problem of the proposed general algorithm for testing the identity of two population's distributions using fuzzy samples is that it works about an order slower than the corresponding algorithm to solve the same problem using two rigid samples¹⁶. In any computer-intensive family of methods (as the Bootstrap procedures) the delay of execution time by such magnitude cannot be disregarded, but should be addressed in future research.

References

1. O. P. Faugeras, Maximal Coupling of Empirical Copulas for Discrete Vectors, *Journal of Multivariate Analysis* **137** (2015) 179–186
2. K. Tenekedjiev, D. Dimitrakiev and N.D. Nikolova, Building Frequentist Distributions of Continuous Random Variables, *Machine Mechanics* **47** (2002) 164–168
3. S. Gao, Y. Zhong and C. Gu, Random Weighting Estimation of Confidence Intervals for Quantiles, *Australian & New Zealand Journal of Statistics* **55(1)** (2013) 43–53
4. G.L. Poe, K.L. Giraud and J.B. Loomis, Computational Methods for Measuring the Difference of Empirical Distributions, *American Journal of Agricultural Economics* **87(2)** (2005) 353–365
5. A.Chernobai, S. T. Rachev and F. Fabozzi, Composite Goodness-of-Fit Tests for Left-Truncated Loss Samples, In C.-F. Lee and J.-C. Lee, *Handbook of Financial Econometrics and Statistics* (2014) 575–596
6. D.F. Groebner, P.W. Shannon, Ph. C. Fly and K.D. Smith, *Business Statistics – A Decision-Making Approach*, Eighth Edition (Prentice Hall, USA, 2011) 770–788
7. W. Böhm and K. Hornik, A Kolmogorov-Smirnov Test for r Samples, *Fundamenta Informaticae - Lattice Path Combinatorics and Applications* **117(1-4)** (2012) 103–125
8. H. Srinivasah, S. Srihari and M. Beal, Signature Verification using Kolmogorov-Smirnov Statistic, *Proc. International Graphonomics Society Conference (IGS)* (2005) 152–156
9. Yu. Lemeshko and A. A. Gorbunova, Application and Power of the Nonparametric Kuiper, Watson, and Zhang Tests of Goodness-of-Fit, *Measurement Techniques* **56(5)** (2013a) 465–475
10. X. Jin, T. W. S. Chow, Y. Sun, J. Shan and B. C. P. Lau, Kuiper Test And Autoregressive Model-Based Approach For Wireless Sensor Network Fault Diagnosis, *Wireless Networks* **21** (2015) 829–839
11. Yu. Lemeshko and A. A. Gorbunova, Application of Nonparametric Kuiper and Watson Tests of Goodness-of-Fit for Composite Hypotheses, *Measurement Techniques* **56(9)** (2013b) 965–973
12. W.H. Press, S. A. Teukolski, W. T. Vetterling and B. P. Flannery, *Numerical Recipes – The Art of Scientific Computing* (Third Edition, Cambridge University Press, 2007) 807–818
13. L.Wasserman, *All of Nonparametric Statistics* (Springer Science+Business Media, USA, 2006) 27–35
14. G. González-Rodríguez, M. Montenegro, A. Colubi and M.A. Gil, Bootstrap Techniques and Fuzzy Random Variables: Synergy in Hypothesis Testing with Fuzzy Data, *Fuzzy Sets and Systems* **157 (19)** (2006) 2608–2613
15. A.R. Henderson, The Bootstrap: A Technique for Data-Driven Statistics. Using Computer-Intensive Analyses to Explore Experimental Data, *Clinica Chimica Acta* **359(1-2)** (2005) 1–26
16. N.D. Nikolova, N. Mihaylova, and K. Tenekedjiev, Statistical Proof of Differences Between Samples using Bootstrap Hypothesis Test, *Science, Labor and Capital* **3-4** (2013) 79–90 (in Bulgarian)
17. N. Karavarsamis, A. Robinson, G. Hepworth, A. Hamilton and G. Heard, Comparison of Four Bootstrap-Based Interval Estimators of Species Occupancy and Detection Probabilities, *Australian & New Zealand Journal of Statistics* **55(3)** (2013) 235–252
18. B. Efron, Second Thoughts on the Bootstrap, *Statistical Science*, **18(2)** (2003) 135–140
19. R. Coppi, M.A. Gil and H.A.L. Kiers, The Fuzzy Approach to Statistical Analysis, *Computational Statistics & Data Analysis* **51(1)** (2006) 1–14

20. M.A. Gil, M. López-Díaz and D.A. Ralescu, Overview on the Development of Fuzzy Random Variables, *Fuzzy Sets and Systems* **157(19)** (2006) 2546–2557
21. Th. Denoeux, Maximum Likelihood Estimation from Fuzzy Data using the EM Algorithm, *Fuzzy Sets and Systems* **183** (2011) 72–91
22. R. Viertl, Univariate Statistical Analysis with Fuzzy Data, *Computational Statistics & Data Analysis*, **51(1)** (2006) 133–147
23. R. Viertl, Testing Hypotheses with Fuzzy Data: The Fuzzy p-Value, *Metrika* **59(1)** (2004) 21–29
24. A. Colubi, Statistical Inference About the Means of Fuzzy Random Variables: Applications to the Analysis of Fuzzy- and Real-Valued Data, *Fuzzy Sets and Systems, Special Issue on Fuzzy Sets, Probabilities, and Statistics: Gaps and Bridges* **160(3)** (2009) 344–356
25. P.G. Panayotov, *Contemporary Methods for Diagnosis and Surgical Treatment of Significant Chronic Ischemic Mitral Regurgitation* (Extended Abstract of PhD thesis, Medical University-Varna, 2013, http://mu-varna.bg/BG/Research/Documents/Doktor/panayotov24.10/avtoreferat_9_EN.pdf) 7–18
26. O. Buchatala and S. Sick, Goodness of Fit: Measures for Fuzzy Classifier, *Proc. IEEE Symposium FOCI 2007* (2007) 201–207
27. A. Kovács, P. Sótónyi, A.I. Nagy, K. Tenekedjiev, N. Wohner, E. Komorowicz, E. Kovács, N. Nikolova, L. Szabó, I. Kovalszky, R. Machovich, Z. Szelid, D. Becker, B. Merkely and K. Kolev, Ultrastructure And Composition of Thrombi in Coronary and Peripheral Artery Disease: Correlations With Clinical and Laboratory Findings, *Thrombosis research* **135(4)** (2015) 760–766
28. N. Wohner, P. Sótónyi, R. Machovich, L. Szabó, L. Tenekedjiev, M.M. Silva, C. Longstaff and K. Kolev, Lytic Resistance of Fibrin Containing Red Blood Cells, *Arteriosclerosis, Thrombosis, and Vascular Biology* **31(10)** (2011) 2306–13
29. D. Henderson, C.F. Parmeter and R.R. Russell, Modes, Weighted Modes, and Calibrated Modes: Evidence of Clustering using Modality Tests, *Journal of Applied Econometrics* **23(5)** (2008) 607–638