



# Kolmogorov-Smirnov Test for Interval Data

Sébastien Destercke, Olivier Strauss

## ► To cite this version:

Sébastien Destercke, Olivier Strauss. Kolmogorov-Smirnov Test for Interval Data. IPMU: Information Processing and Management of Uncertainty in Knowledge-Based, Jul 2014, Montpellier, France. pp.416-425, 10.1007/978-3-319-08852-5\_43 . hal-01045013

**HAL Id: hal-01045013**

**<https://hal.science/hal-01045013>**

Submitted on 24 Jul 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Kolmogorov-Smirnov test for interval data

Sebastien Destercke<sup>1</sup> and Olivier Strauss<sup>2</sup>

<sup>1</sup> Heudiasyc, UMR 7253, Rue Roger Couttolenc, 60203 Compiègne, France  
sebastien.destercke@hds.utc.fr

<sup>2</sup> LIRMM (CNRS & Univ. Montpellier II), 161 rue Ada, F-34392 Montpellier cedex 5, France  
olivier.strauss@lirmm.fr

**Abstract.** In this paper, we are interested in extending the classical Kolmogorov-Smirnov homogeneity test to compare two samples of interval-valued observed measurements. In such a case, the test result is interval-valued, and one major difficulty is to find the bounds of this set. We propose a very efficient computational method for approximating these bounds by using a p-box (pairs of upper and lower cumulative distributions) representation of the samples.

**Keywords:** Interval data, homogeneity test, approximation, p-box

## 1 Introduction

In many applications, the precise value of data may only be known up to some precision, that is it may be interval-valued. Common examples are censored data (e.g., censor limitations) or digital data. When performing statistical tests, ignoring this imprecision may lead to unreliable decisions. For instance, in the case of digital data, quantization can hide the information contained in the data and provide unstable decision.

It is therefore advisable to acknowledge this imprecision in statistical tests, if only to provide results robust to this imprecision. By robust, we understand tests that will remain cautious (i.e., will abstain to say something about the null hypothesis) if not enough information is available. However, treating this imprecision usually leads to an increased computational costs, as shown by various authors in the past [6,7,3]. This means that developing efficient methods to compute statistics with interval data is a critical issue.

In this paper, we explore the extension of the Kolmogorov-Smirnov (KS) homogeneity test to interval data, and more precisely its computational aspects. To our knowledge, this aspect has not been considered in the past, even if some but not much works on the KS test with interval or fuzzy data exist [4,5]. Approximate and exact bounds that are straightforward to compute are provided in Section 3, while notations and reminders are given in Section 2.

In Section 4, we illustrate our results on a image based medical diagnosis problem. Indeed, in such problems a common task is to detect whether two regions of a quantized image have similar pixel distributions.

## 2 Preliminary material

Komogorov-Smirnov (KS) homogeneity test [1] is commonly used to compare two samples  $A = \{a_i | i = 1, \dots, n, a_i \in \mathbb{R}\}$  and  $B = \{b_i | i = 1, \dots, m, b_i \in \mathbb{R}\}$  of measurements to determine whether or not they follow the same probability distribution. Those samples are supposed to be independently drawn from a continuous one-dimensional real-valued probability distributions.

If  $F_A$  ( $F_B$ ) denote the empirical cumulative distributions built from  $A$  ( $B$ ), that is if

$$F_A(x) = \frac{\#\{a \in A | a \leq x\}}{n} \quad (1)$$

with  $\#E$  the cardinal of a set  $E$ , then the KS test statistic  $KS$  is defined by:

$$KS(A, B) = \sup_{x \in \mathbb{R}} |F_A(x) - F_B(x)|$$

Under the null hypothesis  $H_0$  that the two-samples are drawn from the same distribution, the statistic  $\beta(n, m)KS(A, B)$  converges to the Kolmogorov distribution, with  $\beta(n, m) = \sqrt{\frac{1}{n} + \frac{1}{m}}$ . Using the critical values of the Kolmogorov distribution, the null hypothesis can be rejected at level  $\alpha$  if  $KS(A, B) > \beta(n, m)\kappa_\alpha$ . One common value of this rejection threshold is  $\kappa_{0.05} = 1.36$ .

As this test makes very few assumptions about the samples (i.e., it is non-parametric) and aims at testing a complex hypothesis (with respect to, e.g., comparing two means), it requires in practice relatively large samples to properly reject the null hypothesis.

In this paper, we explore the case where observations are interval-valued, i.e., they correspond to two sets  $[A] = \{[a_i, \bar{a}_i] | i = 1, \dots, n\}$  and  $[B] = \{[b_i, \bar{b}_i] | i = 1, \dots, m\}$  of real-valued intervals. As recalled in the introduction and further explored in Section 4, such imprecision may be the result of some quantization process.

In the next section, we study the interval-valued statistic resulting from such data, and in particular provide efficient approximative (and sometimes exact) bounds for it, using the notion of p-box.

## 3 Kolmogorov-Smirnov Test with interval-valued data

Let us first introduce some notations. We will call *selection* of  $[A]$  a set  $S_{[A]}$  of values  $S_{[A]} := \{a_i | i = 1, \dots, n, a_i \in [a_i, \bar{a}_i]\}$  where each  $a_i$  is picked inside the interval  $[a_i, \bar{a}_i]$ ,  $i = 1, \dots, n$ . We will denote by  $\mathcal{S}([A])$  the set of all selections of  $[A]$ . To a selection  $S_{[A]}$  corresponds an empirical cumulative distribution  $F_{S_{[A]}}$  obtained by Eq. (1), and we denote by  $\mathcal{F}([A])$  the (non-convex) set of such empirical cumulative distributions.

Given this, the imprecise Kolmogorov-Smirnov Test

$$[KS]([A], [B]) = [\underline{KS}([A], [B]), \overline{KS}([A], [B])]$$

is an interval such that

$$\underline{KS}([A], [B]) = \inf_{\substack{S_{[A]} \in \mathcal{S}([A]), \\ S_{[B]} \in \mathcal{S}([B])}} \sup_{x \in \mathbb{R}} |F_{S_{[A]}}(x) - F_{S_{[B]}}(x)|, \quad (2)$$

$$\overline{KS}([A], [B]) = \sup_{\substack{S_{[A]} \in \mathcal{S}([A]), \\ S_{[B]} \in \mathcal{S}([B])}} \sup_{x \in \mathbb{R}} |F_{S_{[A]}}(x) - F_{S_{[B]}}(x)|. \quad (3)$$

Computing such values is not, a priori, a trivial task since the number of possible selections for both sets of intervals  $[A]$  and  $[B]$  are usually infinite. It should however be noted that, as the empirical cumulative distributions can only take a finite number of values (i.e.,  $\{0, 1/n, 2/n, \dots, 1\}$  for  $[A]$ ), so does the test. Yet, we are only interested in the extreme values it can take.

In the sequel, we propose to use the formalism of p-boxes to approximate those bounds  $\underline{KS}([A], [B])$  and  $\overline{KS}([A], [B])$

### 3.1 Approximating p-box

A p-box [2]  $[\underline{F}, \overline{F}]$  is a pair of cumulative distributions such that  $\underline{F}(x) \leq \overline{F}(x)$  for any  $x \in \mathbb{R}$ . The usual notion of cumulative distribution is retrieved when  $\underline{F} = \overline{F}$ , and a p-box usually describes an ill-known cumulative distribution that is known to lie between  $\underline{F}$  and  $\overline{F}$ . That is, to a p-box  $[\underline{F}, \overline{F}]$  we can associate a set  $\Phi([\underline{F}, \overline{F}])$  of cumulative distributions such that

$$\Phi([\underline{F}, \overline{F}]) = \{F | \forall x \in \mathbb{R}, \underline{F}(x) \leq F(x) \leq \overline{F}(x)\}.$$

Here, we will use it as an approximating tool.

For a set of intervals  $[A]$ , let us denote by  $S_{\underline{a}}$  and  $S_{\overline{a}}$  the particular selections  $S_{\underline{a}} = \{a_i | i = 1, \dots, n\}$  and  $S_{\overline{a}} = \{\overline{a}_i | i = 1, \dots, n\}$ . Then, we define the p-box  $[\underline{F}_{[A]}, \overline{F}_{[A]}]$  approximating  $[A]$  as

$$\underline{F}_{[A]} := F_{S_{\overline{a}}} \quad \text{and} \quad \overline{F}_{[A]} := F_{S_{\underline{a}}}.$$

We have the following property

**Proposition 1.** *Given a set of intervals  $[A]$ , we have  $\mathcal{F}([A]) \subseteq \Phi([\underline{F}_{[A]}, \overline{F}_{[A]}])$*

*Proof.* Consider a given selection  $S_A$ . For every  $a_i$  in this selection, we have

$$\underline{a}_i \leq a_i \leq \overline{a}_i.$$

Since this is true for every  $i = 1, \dots, n$ , this means that  $F_{S_A}$  is stochastically dominated<sup>3</sup> by  $F_{S_{\overline{a}}}$  and stochastically dominates  $F_{S_{\underline{a}}}$ , i.e.

$$F_{S_{\overline{a}}}(x) \leq F_{S_A}(x) \leq F_{S_{\underline{a}}}(x), \forall x \in \mathbb{R}$$

and as this is true for every selection  $S_A$ , we have  $\mathcal{F}([A]) \subseteq \Phi([\underline{F}_{[A]}, \overline{F}_{[A]}])$ . To see that the inclusion is strict, simply note that  $F_{S_A}$  can only take a finite number of values, while cumulative distributions in  $\Phi([\underline{F}_{[A]}, \overline{F}_{[A]}])$  can be strictly monotonous.

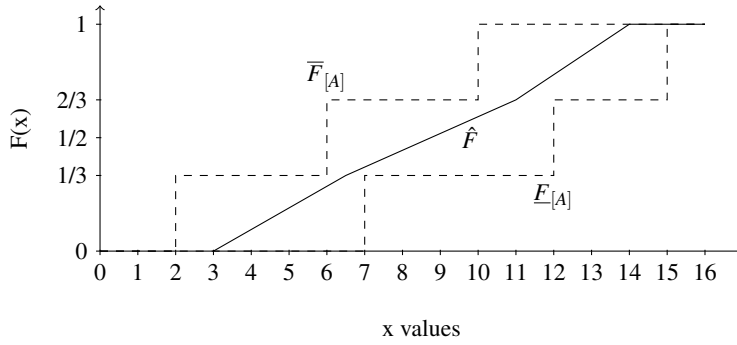
<sup>3</sup> Recall that  $F_1$  stochastically dominates  $F_2$  if  $F_1 \leq F_2$ .

This shows that the associated (convex) set  $\Phi([E_{[A]}, \bar{F}_{[A]}])$  is actually a conservative approximation of  $\mathcal{F}([A])$ . The next example illustrates both the p-box  $[E_{[A]}, \bar{F}_{[A]}]$  and Proposition 1.

*Example 1.* Consider the case where we have 3 sampled intervals, with the three following intervals:

$$\begin{aligned} [\underline{a}_1, \bar{a}_1] &= [2, 7] \\ [\underline{a}_2, \bar{a}_2] &= [6, 12] \\ [\underline{a}_3, \bar{a}_3] &= [10, 16] \end{aligned}$$

Figure 1 illustrates the obtained p-box and one cumulative distribution ( $\hat{F}$ ) included in  $\Phi([E_{[A]}, \bar{F}_{[A]}])$ . However,  $\hat{F}$  is not in  $\mathcal{F}([A])$ , since any empirical cumulative distribution obtained from a selection on 3 intervals can only takes its values in the set  $\{0, 1/3, 2/3, 1\}$ .



**Fig. 1.** P-box of Example 1

### 3.2 Approximating $\widetilde{KS}([A], [B])$ and $\widetilde{KS}([A], [B])$

Consider two samples  $[A]$  and  $[B]$  and the associated p-boxes  $[E_{[A]}, \bar{F}_{[A]}]$  and  $[E_{[B]}, \bar{F}_{[B]}]$ . We can now introduce the approximated imprecise KS Test  $\widetilde{KS} = [\widetilde{KS}, \widetilde{KS}]$  such that:

$$\widetilde{KS}([A], [B]) = \sup_{x \in \mathbb{R}} \max\{|\bar{F}_{[A]}(x) - E_{[B]}(x)|, |E_{[A]}(x) - \bar{F}_{[B]}(x)|\}, \quad (4)$$

$$\widetilde{KS}([A], [B]) = \sup_{x \in \mathbb{R}} D_{[A], [B]}(x), \quad (5)$$

with

$$D_{[A], [B]}(x) = \begin{cases} 0 & \text{if } [E_{[A]}(x), \bar{F}_{[A]}(x)] \cap [E_{[B]}(x), \bar{F}_{[B]}(x)] \neq \emptyset \\ \min\{|\bar{F}_{[A]}(x) - E_{[B]}(x)|, |E_{[A]}(x) - \bar{F}_{[B]}(x)|\} & \text{otherwise} \end{cases}$$

These approximations are straightforward to compute (if  $n + m$  intervals are observed, at worst they require  $2n + 2m$  computations once the p-boxes are built). We also have the following properties:

**Proposition 2.** *Given a set of intervals  $[A]$  and  $[B]$ , we have  $\widetilde{KS}([A], [B]) = \overline{KS}([A], [B])$*

*Proof.* The value  $\widetilde{KS}([A], [B])$  is reached on  $x$  either for a pair  $\{\overline{F}_{[A]}, \underline{F}_{[B]}\}$  or  $\{\underline{F}_{[A]}, \overline{F}_{[B]}\}$ . As any pair  $F_1, F_2$  with  $F_1 \in \Phi(\underline{F}_{[A]}, \overline{F}_{[A]})$  and  $F_2 \in \Phi(\underline{F}_{[B]}, \overline{F}_{[B]})$  would have a KS statistic lower than  $\widetilde{KS}([A], [B])$ , and given the inclusion of Proposition 1, this means that  $\overline{KS}([A], [B]) \leq \widetilde{KS}([A], [B])$ . To show that they coincide, it is sufficient to note that all distributions  $\underline{F}_{[A]}, \overline{F}_{[A]}, \underline{F}_{[B]}, \overline{F}_{[B]}$  can be obtained by specific selections (i.e., the one used to build the p-boxes). ■

This shows that the upper bound is exact. Concerning the lower bound, we only have the following inequality:

**Proposition 3.** *Given a set of intervals  $[A]$  and  $[B]$ , we have  $\underline{KS}([A], [B]) \leq \underline{KS}([A], [B])$*

*Proof.* Immediate, given the inclusion of Proposition 1 and the fact that  $\underline{KS}([A], [B])$  is the minimal KS statistics reached by a couple of cumulative distributions respectively in  $[\underline{F}_{[A]}, \overline{F}_{[A]}]$  and  $[\underline{F}_{[B]}, \overline{F}_{[B]}]$  ■

And unfortunately this inequality will usually be strict, as shows the next example.

*Example 2.* Consider the case where  $n = 2, m = 3$  and where  $\bigcap_{i=1}^n [a_i, \bar{a}_i] = \emptyset, \bigcap_{i=1}^m [b_i, \bar{b}_i] = \emptyset$ . This means that, for every selection  $S_{[A]} \in \mathcal{S}([A])$  and  $S_{[B]} \in \mathcal{S}([B])$ , we have that the empirical cumulative distributions  $F_{S_{[A]}}$  and  $F_{S_{[B]}}$  respectively takes at least one value in  $\{1/2\}$  and in  $\{1/3, 2/3\}$ . This means that  $\underline{KS}([A], [B]) \neq 0$  (as every cumulative distributions coming from selections will assume different values), while it is possible in such a situation to have  $\underline{KS}([A], [B]) = 0$ .

Consider the following example:

$$\begin{array}{ll} [a_1, \bar{a}_1] = [1, 8] & [b_1, \bar{b}_1] = [2, 7] \\ [a_2, \bar{a}_2] = [9, 15] & [b_2, \bar{b}_2] = [6, 12] \\ & [b_3, \bar{b}_3] = [10, 16] \end{array}$$

A simple look at Figure 2 allows us to see that  $\underline{KS}([A], [B]) = 0$  in this case.

The inequality between  $\underline{KS}([A], [B])$  and  $\underline{KS}([A], [B])$  can also be strict when  $\underline{KS}([A], [B]) \neq 0$ . It should be noted that the discrepancy between  $\underline{KS}([A], [B])$  and  $\underline{KS}([A], [B])$  will decrease as the number of sampled intervals increases. Finally, a noticeable situation where  $\underline{KS}([A], [B])$  will be an exact bound ( $\underline{KS}([A], [B]) = \underline{KS}([A], [B])$ ) is when  $[\underline{F}_{[A]}, \overline{F}_{[A]}]$  and  $[\underline{F}_{[B]}, \overline{F}_{[B]}]$  are disjoint, that is either  $\underline{F}_{[A]} > \overline{F}_{[B]}$  or  $\overline{F}_{[A]} < \underline{F}_{[B]}$ .

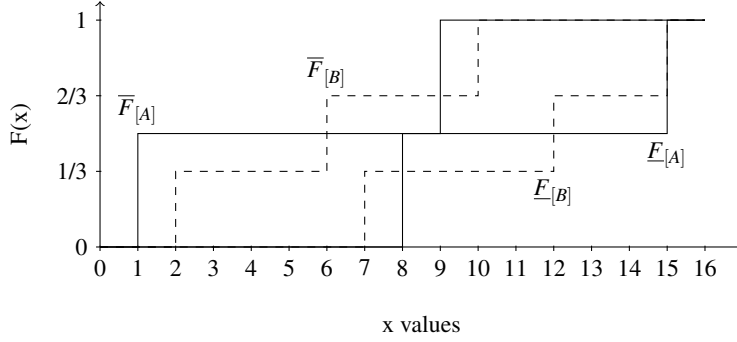


Fig. 2. P-boxes of Example 2

### 3.3 Decision making using an imprecise-valued test

One of the main features of this extension is that it provides a pair of (conservative) bounds  $\underline{KS}([A], [B])$  and  $\widetilde{KS}([A], [B])$  rather than a precise value  $KS(A, B)$ . In contrast with usual tests that either reject or do not reject an hypothesis, this leads to three possible decisions: the answer to the test can be *yes*, *no* or *unknown*, the last one occurring when available information is insufficient.

In fact, interpreting this test is straightforward. Let  $\gamma = \beta(n, m)\kappa_\alpha$  be the significance level.

- If  $\underline{KS}([A], [B]) > \gamma$  then we can conclude that there is no possible selections  $S_{[A]}$  of  $[A]$  and  $S_{[B]}$  of  $[B]$  such that  $KS(S_{[A]}, S_{[B]}) \leq \gamma$  and thus the hypothesis that the two-samples are drawn from the same distribution can be rejected at a level  $\alpha$ .
- On the contrary, if  $\widetilde{KS}([A], [B]) < \gamma$  then there is no possible selections  $S_{[A]}$  of  $[A]$  and  $S_{[B]}$  of  $[B]$  such that  $KS(S_{[A]}, S_{[B]}) \geq \gamma$  and thus the hypothesis that the two-samples are drawn from the same distribution cannot be rejected at a level  $\alpha$ .
- Otherwise, we will conclude that our information is too imprecise to lead to a clear decision about rejection.

This new test will therefore point out those cases where the data imprecision is too important to lead to a clear decision. As we shall see in the next section, it allows one to deal with quantization in a new way, namely it can detect when the disturbance or loss of information induced by the quantization makes the test inconclusive. It should be noted that, as  $\widetilde{KS}([A], [B])$  is an approximated lower bound, indecision may also be due to this approximation, yet experiments of the next section indicate that this approximation is reasonable.

## 4 Experimentation

The experimentation we propose is based on a set of medical images acquired by a gamma camera. In such applications, statistical hypothesis testing is often used to determine whether pixels distribution in two different regions of an image are similar or

not. Physicians usually try to control the probability of making a decision leading to harmful consequences, and make it as low as possible (usually 0.05).

The advantage of using a KS test in this case is that it makes very few assumption about the distribution. However, in such applications, it is quite common to deal with quantized information, i.e., real-valued information constrained to belong to a small subset of (integer) values. Since the KS test is designed to compare pairs of continuous distributions, it is necessary to ensure that the statistical test is robust with respect to the data model. Indeed, the value of the statistic computed from quantized data may differ markedly from the calculation based on original (non-quantized) but unavailable data.

Physicians would usually try to avoid a wrong decision, and prefer to acquire additional data when the actual data are not fully reliable. Thus, knowing that no decision can be taken based on the current set of data is a valuable piece of information.

We illustrate this weakness of the usual KS test with a set of medical images acquired by a gamma camera (nuclear medicine images) whose values are quantified on a restricted number of values. This experiment also highlights the ability of the extended KS test to avoid wrong decisions induced by quantization. It aims at mimicking real medical situations where the nuclear physician has to compare the distribution of values in two regions of interest in order to decide whether or not a patient has a specific disease.

The set of images is made of 1000 planar acquisitions of a Hoffman 2-D brain phantom (acquisition time: 1 second; average count per image 1.5 kcounts,  $128 \times 128$  images to satisfy the Shannon condition), representing 1000 measures of a random 2D image (see Figure (3)). Due to the fact that nuclear images are obtained by counting the photons that have been emitted in a particular direction, pixel values in a nuclear image can be supposed to be contaminated by Poisson distributed noise. Due to the very short acquisition time, the images were very noisy, i.e. the signal to noise ratio was very low. More precisely, the average pixel value in the brain corresponded to a 69% coefficient of variation of the Poisson noise. Moreover, the number of different possible values to be assigned to a pixel was low and thus, within those images, the impact of quantization was high: pixel possible values were  $\{0, 256, 512, 768, 1024, 1280, 1536, 1792, 2048\}$ .

To obtain less noisy and less quantized images, we summed the raw images (see e.g. Figure (4)). The higher the number of summed images, the higher the average pixel value, and thus the higher the signal to noise ratio and the higher is the number of possible values for each pixel. When summing the 1000 raw images, we obtained the high dynamic resolution and high signal to noise ratio image depicted in Figure (5).a.

We use the KS test to decide whether the two regions depicted in Figures (5).b and (5).c can be considered as being similar or not (the null hypothesis). Considering the number of pixels in each region ( $n = 183$ ,  $m = 226$ ), the significance level for a p-value  $\alpha = 0.05$  is  $\gamma \approx 0.1910$ . Testing the two regions with the reference image (Figure (5).a) provides the following values:  $KS(A, B) \approx 0.2549$ ,  $\underline{KS}([A], [B]) \approx 0.2505$ ,  $\widetilde{KS}([A], [B]) \approx 0.2549$ , leading to conclude that the similarity of regions A and B should be rejected at a level 0.05, which can be considered as our ground truth.

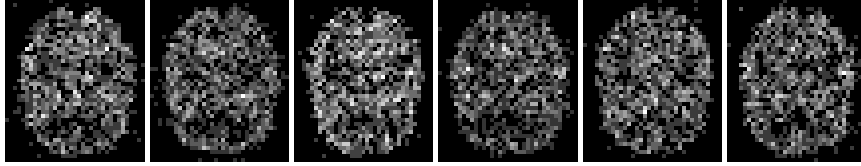
We use the KS test for comparing the same regions but with 300 pairs of images that have been randomly selected in the set of 1000 original images. In that case, the classical test accepts the similarity of the two regions, while the imprecise test is inconclusive



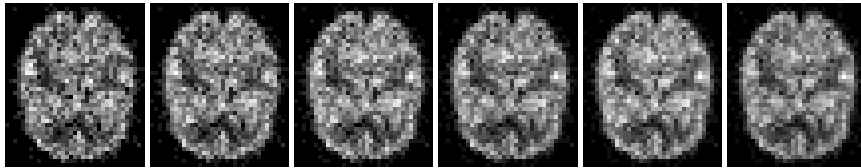
for each pairs:  $\underline{KS}([A], [B]) < \gamma < \widetilde{KS}([A], [B])$ . We now do the same test with images having a higher dynamic obtained by summing  $p = 2, 3, \dots, 40$  images. For each value of  $p$ , we count the number of times the classical test provides the right answer, i.e. reject the null hypothesis at level 0.05 ( $\gamma \leq KS(A, B)$ ). We then compute the ratio of this count over 300. For the extended KS test, we compute two ratios: the ratio of times when  $\gamma \leq \underline{KS}([A], [B])$ , i.e. we **must** reject the null hypothesis at level 0.05, and the ratio of times when  $\gamma \leq \widetilde{KS}([A], [B])$ , i.e. we **can** reject the null hypothesis at level 0.05. We also compute the number of times where the test is inconclusive, i.e.  $\underline{KS}([A], [B]) < \gamma < \widetilde{KS}([A], [B])$ .

Figure (6) plots these ratio versus  $p$ , the number of summed images. On one hand, concerning the classical KS test, it can be noticed that depending on the quantification level, the answer to the test differs. In fact, when the number of pixel's possible values is low, the test concludes that  $H_0$  cannot be rejected most of the time, leading to a decision that the two distributions are similar even though they are not. When  $p$  increases, so increases pixel's possible values and increases the ratio of correct answer. Thus, quantization has a high impact on the conclusions of a classical KS test.

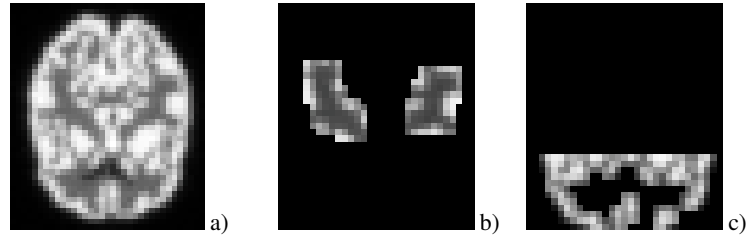
On the other hand, concerning the extended KS test, it can be noticed that the null hypothesis **can** always be rejected. The impact of the quantization only affects the ratio of times when the null hypothesis **must** be rejected. Thus the impact of quantization here is much more sensible, in the sense that when quantization is too severe (information is too poor), the test abstains to make a decision. Also, in all cases, the test is either inconclusive or provides the right answer, and is therefore never wrong, which is what we could expect from a robust test.



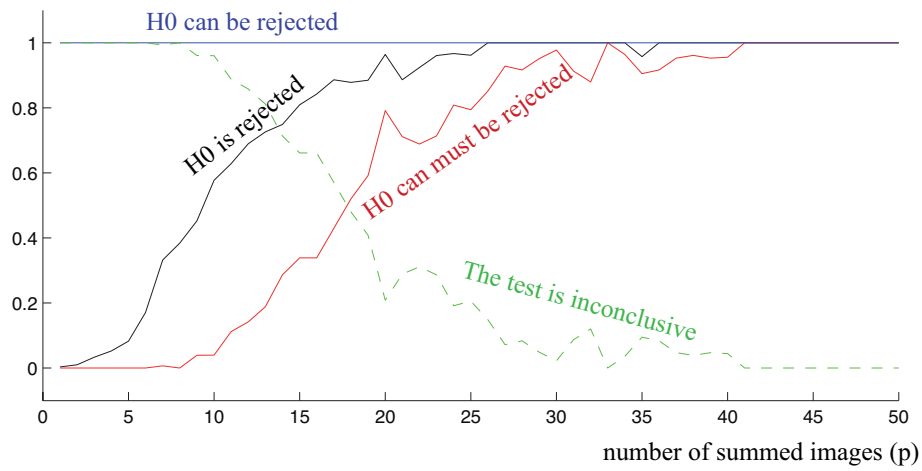
**Fig. 3.** 6 acquisitions of the Hoffman 2-D brain phantom



**Fig. 4.** 6 images obtained by summing up 10 raw acquisitions of the Hoffman 2-D brain phantom



**Fig. 5.** Reference image obtained by summing the 1000 raw images (a), region A (b) and region B (c) selected on the reference image.



**Fig. 6.** Correct decision ratio with the classical test (black) and with the extended test (blue for *can be rejected*, red for *must be rejected*), superimposed with the proportion of times extended test is inconclusive (green).

## 5 Conclusions

In this paper, we have introduced efficient methods to approximate the bounds of the KS test with interval-valued data. We have demonstrated that the upper bound is exact while the lower bound is, in general, only a lower approximation. However, the experiments have shown that this is not too conservative approximation and still allows to take decision when enough information is available.

The obvious advantages of this paper proposal is its efficiency (computational time is almost linear in the number of sampled intervals), however we may search in the future for exact rather than approximated lower bounds. Since KS test result only depends on the ordering (i.e., ranking) of sampled elements between them, a solution would be to explore the number of possible orderings among elements of  $[A]$  and  $[B]$ , or to identify the orderings for which the lower bound is obtained (the number of such orderings, while finite, may be huge).

Finally, it would also be interesting to investigate other non-parametric homogeneous tests, such as the Cramer-Von Mises one.

## Acknowledgements

This work was partially carried out in the framework of the Labex MS2T, which was funded by the French Government, through the program "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02)

## References

1. W. J. Conover. *Practical non-parametric statistic*. Wiley, New York, 3rd edition, 1999.
2. S. Ferson, L. Ginzburg, V. Kreinovich, D. M. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.
3. P.-A. Hébert, M.-H. Masson, and T. Denoeux. Fuzzy multidimensional scaling. *Computational Statistics & Data Analysis*, 51(1):335–359, 2006.
4. G. Hesamian and J. Chachi. Two-sample Kolmogorov–Smirnov fuzzy test for fuzzy random variables. *Statistical Papers*, page (To appear), 2014.
5. C. Mballo and E. Diday. Decision trees on interval valued variables. *The electronic journal of symbolic data analysis*, 3:8–18, 2005.
6. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang. *Computing Statistics Under Interval and Fuzzy Uncertainty: Applications to Computer Science and Engineering*. Springer Publishing Company, Incorporated, 2011.
7. J. Otero, L. Sánchez, I. Couso, and A. M. Palacios. Bootstrap analysis of multiple repetitions of experiments using an interval-valued multiple comparison procedure. *J. Comput. Syst. Sci.*, 80(1):88–100, 2014.