Comparing Interval-Valued Estimations with Point-Valued Estimations

Hugo Saulnier, Olivier Strauss, Ines Couso

▶ **To cite this version:**

Hugo Saulnier, Olivier Strauss, Ines Couso. Comparing Interval-Valued Estimations with Point-Valued Estimations. IPMU 2016 - 16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Jun 2016, Eindhoven, Netherlands. pp.595-604, 10.1007/978-3-319-40581-0_48 . lirmm-01488052

HAL Id: lirmm-01488052

https://hal-lirmm.ccsd.cnrs.fr/lirmm-01488052

Submitted on 17 May 2022

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.

Comparing Interval-Valued Estimations with Point-Valued Estimations

Hugo Saulnier[1,2(✉)], Olivier Strauss[1], and Ines Couso[2]

[1] LIRMM, 161 Rue Ada, Montpellier, France
{hsaulnier,strauss}@lirmm.fr
[2] Universidad de Oviedo, Oviedo, Spain
couso@uniovi.es

Abstract. In the last decade, numerous proposals have been made to deal with imprecision in estimation problems. Those approaches, many of which involve dealing with interval-valued outputs, deal with the subtle difference between uncertainty and imprecision. One of the crucial points − which to our knowledge has never been addressed − is "how to compare an interval-valued method with a precise valued method?"

The usual way to compare two estimation methods is to use benchmark data with ground truths and to compute a distance between the estimates of each method and the ground truth. However, most of the mathematical available extensions of distances are either biased in favor of a precise approach or in favor of an imprecise approach.

This paper proposes a new tool, the weighted variation of the midpoint distance (WVD), that is more suitable to achieve this kind of comparison, dealing with imprecision with a particular semantic. After reviewing existing distances, we introduce the WVD, first from an intuitive perspective, then from a more mathematical point of view. Its very satisfactory properties are highlighted through an experiment.

Keywords: Interval-valued data · Imprecise probabilities · Engineering

1 Introduction

Scientists willing to consider imprecise data and methods in their analysis (such as [1,2] or [3]) face the problem of comparing methods with interval-valued output with methods with precisely valued outputs. In this paper, we consider imprecise valued regression methods producing interval valued estimates of a precise reference. One of the most usual ways for assessing a preference between one method and another is to use a set of benchmark data to compute a distance between the output of each method and the known ground truth, the current fashion in engineering problems being to rather use the $\mathcal{L}_1$ distance (see [4] for example). However, if comparing two interval-valued methods or two precise valued methods is straightforward, a comparison between an interval-valued method and a precise valued method is more intricate since any existing extension of the $\mathcal{L}_1$ distance is either biased towards or against imprecision: the

supremum (or Hausdorff[1]) distance tends to disfavor imprecise valued estimates while the *infimum* distance promotes immoderately imprecision.

In order to establish a preference between methods we need to compute the distance between the reference (precise) and the estimate (precise or imprecise). This preference must express as wisely as possible the intricate pros and cons of using an imprecise estimation versus using a precise estimation. Figure 1 illustrates the kind of situation we could fall into. A reference (plain line) is estimated (dotted line) in four different settings.

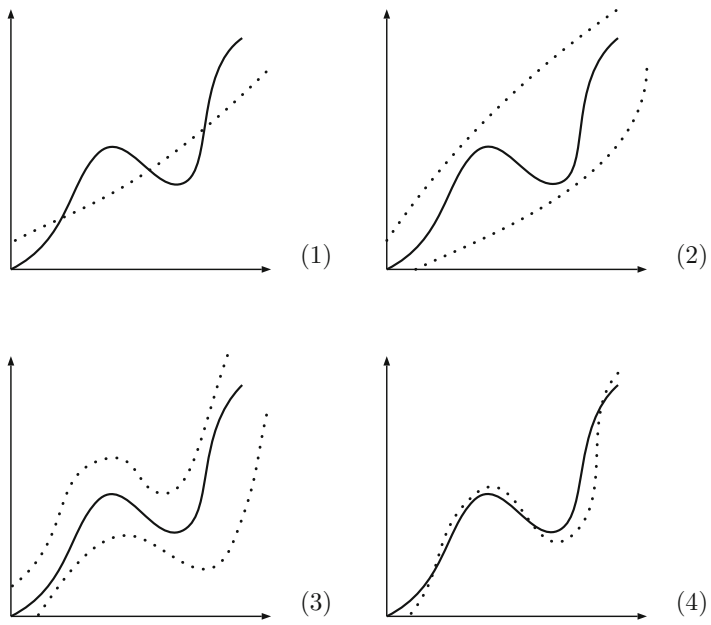

Fig. 1. Precise estimations (1,4) and imprecise estimation (2,3). Plain line is the reference, dotted lines are estimations.

Probably (2) would be preferred to (1). The estimation of (2) is imprecise but informative. It describes the possible variations of the reference. On the other hand the precise estimation of (1) gives an inaccurate description of the data. Again (3) is better than (2), because it has narrower bounds while still being informative. But (4) should be preferred among all, because it is both precise and accurate.

In this paper, we propose a nice candidate to achieve this kind of ordering. The rest of the paper will be organized in three main sections: after reviewing existing metrics, we introduce the WVD, first from an intuitive perspective, then

[1] The Hausdorff distance is also called the Pompeiu-Hausdorff distance. Here we simply refer to it as the Hausdorff distance.

from a more mathematical point of view. Finally some experiments highlight its very nice properties.

2 A Review of Existing Distances Between Points and Intervals

2.1 Notations

Let $\mathbb{R}$ be the real line and $\mathbb{IR}$ be the set of closed intervals of $\mathbb{R}$. X will denote a closed interval of $\mathbb{IR}$, $\underline{x} \in \mathbb{R}$ its lower element, $\overline{x} \in \mathbb{R}$ its upper element, $\tilde{x} = \frac{\underline{x}+\overline{x}}{2}$ its center, and $r = \frac{\overline{x}+\underline{x}}{2}$ its radius.

Here we are interested in defining distances between a finite sequence of points (scalar vector) and a finite sequence of intervals (interval vector). In the following we will consider distances from a reference vector of scalars $\mathbf{s} = (s_1, \ldots, s_N)$ to an indexed collection (vector) of intervals $\mathbf{X} = (X_1, \ldots, X_N)$ where $\forall i \in \{1, \ldots, N\}, X_i = [\underline{x}_i, \overline{x}_i]$. We also denote by $\tilde{\mathbf{x}}$ the vector of the center values of $\mathbf{X}$. The literature includes different alternatives to compute the distance between pairs of intervals of the real line. Based on each of those proposals, one can define the distance from a point $s \in \mathbb{R}$ to an interval $X \in \mathbb{IR}$ as the distance between the singleton $\{s\}$ (which is in turn an element in $\mathbb{IR}$) and X. We can therefore define the distance from $\mathbf{s} = (s_1, \ldots, s_N)$ to $\mathbf{X}$ as the arithmetic mean of the distances between their components. We next review different proposals of distances between pairs of real intervals from the literature, and construct their associated distances from vectors of points to vectors of intervals.

2.2 Hausdorff Distance

Let (U, d) be a complete metric space. Let $\mathcal{K}(U)$ denote the family of non-empty compact subsets of U. The Hausdorff distance between two non-empty compact subsets $A, B \in \mathcal{K}(U)$ is

$$d_H(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\}.$$

In particular, we can define the *Hausdorff distance from $s \in \mathbb{R}$ to $X \in \mathbb{IR}$* as the Hausdorff distance[2] between $\{s\}$ and X, i.e.:

$$d_H(s, I) = \max\{\sup_{x \in I} d(s, x), \inf_{x \in I} d(s, x)\} = \sup_{x \in I} d(s, x), \tag{1}$$

The Hausdorff distance is one the most widely used distances from points to sets [5]. If d is the $\mathcal{L}_1$ distance, Eq. (1) can be simplified in:

$$d_H(s, I) = |s - \tilde{x}| + r, \tag{2}$$

[2] We will use the same notation d_H in order to denote the mapping defined on $\mathbb{R} \times \mathbb{IR}$ derived from $d_H : \mathcal{K}(\mathbb{R}) \times \mathcal{K}(\mathbb{R}) \to \mathbb{R}$. Obviously this new mapping does not satisfy metric properties. It is not even applied to objects of the same kind.

where $\tilde{x}$ and r respectively denote the mid-point and the radius of the closed interval $X = [\tilde{x}-r, \tilde{x}+r]$. Computation of the distance from $\mathbf{s} \in \mathbb{R}^N$ to $\mathbf{X} \in \mathbb{IR}^N$ is obtained by averaging Expression (2) on their components:

$$d_H(\mathbf{s}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} (|s_i - \tilde{x}_i| + r_i). \tag{3}$$

2.3 Mid-Spread Distance

Expression (2) considers the interval X as a pair composed of its centre $\tilde{x}$ and its radius r like in the mid-spread distance approach [6], where the distance is computed as a linear combination of the distance between the two centers and the distance between the two radii (the radius of s being equal to 0). The mid-spread distance between two intervals $X = [\tilde{x} - r, \tilde{x} + r]$ and $Y = [\tilde{y} - r', \tilde{y} + r']$ is defined as:

$$d^{\gamma}_{ms}(X, Y) = |\tilde{x} - \tilde{y}| + \gamma |r - r'|.$$

According to it, we can define the *mid-spread distance from s to* $X = [\tilde{x}+r, \tilde{x}+r]$ *as:*

$$d^{\gamma}_{ms}(s, X) = |s - \tilde{x}| + \gamma r.$$

Based on it we can calculate the mid-spread distance from $\mathbf{s}$ to $\mathbf{X}$ as:

$$d^{\gamma}_{ms}(\mathbf{s}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} (|s_i - \tilde{x}_i| + \gamma r_i), \tag{4}$$

where $\gamma \in \mathbb{R}^+$ is a weight given to the radius w.r.t. the mid. This approach can be seen as a generalization of the Hausdorff distance since $d_H = d^1_{ms}$.

2.4 Mid-Point Distance

The mid-point distance leads to a less technical distance. It is just the $\mathcal{L}_1$ distance between $\mathbf{s}$ and $\tilde{\mathbf{x}}$. Within this approach the intervals are reduced to points and the imprecision is simply ignored. This approach is easy both conceptually and computationally. For this reason it is used a lot in practice [7]. It is defined as:

$$d_m(\mathbf{s}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} |s_i - \tilde{x}_i|. \tag{5}$$

It can also be considered either as a special case of the mid-spread distance, where $\gamma = 0$ or as a two step Hausdorff distance. In this last interpretation, the method consists of choosing, as a representative single point of the interval $\mathbf{X}$, the precise value that minimizes the Hausdorff distance of the $\mathcal{L}_1$ distance, i.e. its center $\tilde{\mathbf{x}}$, and then compute the usual distance between $\mathbf{s}$ and this representative.

2.5 Infimum Distance

The classical way to define the distance from a point s to a set X in topology is to consider the shortest distance from s to any point in X, i.e.:

$$d_{\dot{inf}}(s, X) = \inf_{x \in X} d(s, x). \tag{6}$$

When, in particular, d denotes the $\mathcal{L}_1$ distance from a point s to a closed interval $X = [\tilde{x} - r, \tilde{x} + r]$, Expression (7) reduces to:

$$d_{\dot{inf}}(s, X) = \begin{cases} 0 & \text{if} \quad x \in X \\ |\tilde{x} - s| - r & \text{if} \quad x \notin X \end{cases} \tag{7}$$

Based on this definition, we can compute the distance from $\mathbf{s}$ to $\mathbf{X}$ by averaging expression (7) on the N components:

$$d_{\dot{inf}}(\mathbf{s}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} d_{\dot{inf}}(s_i, X_i). \tag{8}$$

This definition captures the notion of imprecision, but does not have enough separating power. As a long as a point is included in the interval, the distance will be zero. It means that an estimation consisting of intervals of the form $(-\infty, +\infty)$, i.e. a completely vacuous estimation, would always have a distance of zero to any precise estimation, because it contains them all. This distance has exactly the opposite flaw than the Hausdorff distance: the wider the interval, the lower the distance. Thus an interval-valued estimation will always be considered as less distant from a reference value than an equivalent precise-valued estimation: $\forall \mathbf{y} \in \mathbf{X}$, $d(\mathbf{s}, \mathbf{y}) \geq d_{\dot{inf}}(\mathbf{s}, \mathbf{X})$.

3 Weighted Variation of the Mid-Point Distance

3.1 Definition

A perfect candidate would be an extension which would not penalize imprecision when it conveys information, for example when it reflects coherently the variability of the quantities under consideration. This is in line with the guaranteed approach of [8] and the confidence interval interpretation of imprecision [9].

The weighted variation of the mid-point distance is simply defined as the mean of deviations from the center of the intervals weighted by the inverse of the radius of the interval:

$$d_w(\mathbf{s}, \mathbf{X}) = \frac{1}{N} \sum_{i=1}^{N} \frac{\overline{r}}{r_i} |s_i - \tilde{x}_i|, \tag{9}$$

with

$$\overline{r} = \frac{1}{N} \sum_{i=1}^{N} r_i.$$

It can be seen as a mid-point distance computed on a space where the imprecision is uniform (usually unitary). It is then back-projected to the original space by multiplying the result by the mean radius of the intervals. The weights of the weighted extension are proportional to the inverses of the radii of the intervals. Thus for a certain index, the wider the radius, the lower will be the impact of an important deviation from the center of the interval. It is a very straightforward way of translating the idea that if an interval estimation suitably describes the variability of the reference, then it should be less penalized. Still, having wide radii will be penalized by increasing the mean radius $\overline{r}$.

Obviously the WVD does not satisfy metric properties. The main problem is that if the vector $\mathbf{X}$ contains both intervals of strictly positive radii and degenerated intervals (i.e. singletons) then the WVD we propose is not defined. It thus does not formally stand as an extension of the $\mathcal{L}_1$ distance. However we shall insist that the aim of this tool is to propose a practical solution to the unsolved problem of computing distances between points and intervals. When we restrict its use to cases where $\mathbf{X}$ contains only non-degenerated intervals − which happens most of the time in practical cases − then the WVD fulfills its role in a way that is both very simple computationally and that makes sense from a formal point of view.

3.2 Formal Interpretation

Let us consider a regression problem where $\mathcal{X} : \Omega \to \mathbb{R}^n$ denotes the vector of attributes and $\mathcal{Y} : \Omega \to \mathbb{R}$ represents the response variable. Let us consider a sample of size N $((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N))$. Let us consider a regression model $f : \mathbb{R}^N \to \mathbb{R}$ and let $\hat{y}_i = f(\mathbf{x}_i)$ be the punctual estimation of y_i based on it. The average distance $\overline{d} = \frac{1}{N}|y_i - \hat{y}_i|$ can be used to estimate the degree of goodness of our model, in terms of $\mathcal{L}_1$ loss function. Let now $X_{i,\alpha} = [\hat{y}_i - c_\alpha r_i, \hat{y}_i + c_\alpha r_i]$ represent a prediction interval, with (exact) confidence level $1 - \alpha$, based on $\mathbf{x}_i$ for the value of the response variable of another individual, randomly picked in the subpopulation of those whose vector of attributes coincides with the vector of attributes of the i^{th} individual of our initial sample, $\mathbf{x}_i$. Due to the variability of this subpopulation, such a response value does not necessarily coincide with the observed y_i. The length of prediction intervals in linear regression problems, under the usual normality assumptions, takes this form of $c_\alpha r_i$. The average radius of the N prediction intervals calculated from the sample is $c_\alpha \overline{r}$. Let us now consider another test sample of size N of pairs of the form $(\mathbf{x}_i, y'_i)$. An unbiased estimation of $1 - \alpha$ based on this test sample is:

$$1 - \hat{\alpha} = \frac{1}{N} \#\{i \,|\, |\hat{y}_i - y'_i| \leq c_\alpha r_i\} = \frac{1}{N} \#\{i \,|\, \frac{|\hat{y}_i - y'_i|}{r_i} \leq c_\alpha\}.$$

Let the random variable $\overline{\mathcal{D}}$ denote the average $\mathcal{L}_1$ distance between the true values y'_i and their estimations $\hat{y}_i$ based on the $\mathbf{x}_i$'s. Let us prove that the WVD from the vector of responses y'_i to the sequence of confidence intervals $X_{i,\alpha}$ is an unbiased estimation of the expectation of $\overline{\mathcal{D}}$. According to above assumptions, for every $j = 1, \ldots, N$, the probability that D_j is less than $c_\alpha \cdot r_j$ is $1-\alpha$. Furthermore, as we have shown, the relative cumulative frequency $\frac{1}{N}\#\{i|\frac{|y'_i-\hat{y}_i|}{r_i} \leq c_\alpha\}$ is an unbiased estimation of $1-\alpha$. In other words, $\hat{F}(c_\alpha) = \frac{1}{N}\#\{i|\frac{|y'_i-\hat{y}_i|}{r_i} \leq c_\alpha\}$ is an unbiased estimation of the cumulative probability $F_{\frac{D_j}{r_j}}(c_\alpha)$. Therefore the relative frequency $\hat{p}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\frac{|y'_i-\hat{y}_i|}{r_i}}(x)$ is an unbiased estimation of the probability $P\left(\frac{D_j}{r_j} = x\right) = P(D_j = r_j \cdot x)$. Thus, the expectation of D_j is unbiasedly estimated by $\hat{E}(D_j) = \frac{1}{N}\sum_{i=1}^{N} \frac{|y'_i-\hat{y}_i|}{r_i} r_j$. Therefore, the expectation of $\overline{\mathcal{D}}$, $E(\overline{D}) = \frac{1}{N}\sum_{j=1}^{N} E(D_j)$ can be unbiasedly estimated by:

$$\hat{E}(\overline{\mathcal{D}}) = \frac{1}{N}\sum_{i=1}^{N} \frac{|y'_i - \hat{y}_i|}{r_i}\overline{r}.$$

4 Experiment

4.1 Experimental Procedure

This section aims at comparing the behavior of the proposed WVD of the $\mathcal{L}_1$ distance to the four other extensions we have presented in Sect. 2. The experiment is based on a data set composed of $K = 3900$ 8×8 subsampled patches extracted from a 2117×3006 high resolution image of the painting *La Joconde* by Leonardo Da Vinci. The reduction factor is set to 5. The subsampling procedures aims at simulating acquisitions of images of this painting using different imagers having the same numerical resolution but different point spread functions.

To do so, we used the so-called imprecise filtering sub-sampling method [10]. It consists of replacing the smoothing anti-aliasing kernel used to transform a high resolution image into a low resolution image by a capacity that represents a convex set of bell-shaped smoothing kernels. Filtering a patch with this method leads to an interval-valued subsampled patch that represent the convex set of all the patches that would have been obtained by subsampling the original patch with all the kernels belonging to the so-defined convex set (here a set of unimodal centered smooth kernels whose support is lower than 8). Practically, each interval-valued patch is composed of an upper patch $\overline{P}_k$ and a lower patch $\underline{P}_k$ (for the k^{th} patch).

4.2 How Consistent Is the Weighted Variation of the Mid-Point Distance?

This experiment aims at illustrating the fact that the extension we propose behaves consistently. Ideally an imprecise patch contains the information of several precise patches. Therefore a distance between imprecise and precise patches which wisely uses the information of imprecision would behave the same way as the $\mathcal{L}_1$ distance would when used to compare precise patches drawn randomly inside the imprecise patches. The comparison is achieved by computing a ratio of imprecise-to-precise $\mathcal{L}_1$ distances on various configuration of patches. If the distribution of this ratio is centered around 1 for a large number of samples, it means that the considered extension is consistent, because it reflects the information conveyed by the imprecise patches in a way that is consistent with how the $\mathcal{L}_1$ distance would behave on a set of precise patches. Moreover, the lower the variance, the more consistent is the extension.

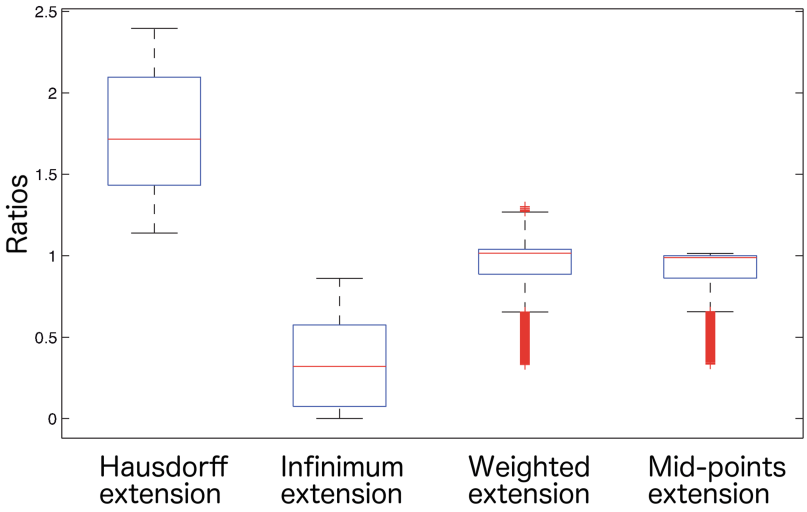

Fig. 2. Box-plot results of the first experiment computed on the four extensions. Weighted extension is WVD.

Table 1. Moments of the ratios.

	Hausdorff	Infinimum	Mid-points	Weighted
mean	1.54	0.503	0.95	0.99
median	1.71	0.32	0.99	1.01
standard deviation	0.36	0.27	0.14	0.16

The experiment is carried as follows. We divide the set of 3900 patches into two subsets of 1950 patches. The motivation for separating our data into two subsets is to compare patches having non-null distances (even if some patches bear some similarities).

We take the first subset ($k = 1, \ldots, 1950$) to define references by considering the mid patches $\tilde{P}_k = \frac{1}{2}(\overline{P}_k + \underline{P}_k)$. We take the second subset ($j = 1951, \ldots, 3900$) to define the imprecise estimations defined by their upper ($\overline{P}_j$) and lower ($\underline{P}_j$) patches. For each imprecise patch of the second subset, we draw 300 precise patches P_j^n ($n = 1, \ldots, 300$) included in the imprecise patch $[\underline{P}_j, \overline{P}_j]$ (300 offers a good tradeoff between a statistically meaningful sampling of the interval and a reasonable computation time). We then compute, for each $(k, j, n) \in [1, 1950] \times [1951, 3900] \times [1, 300]$, a ratio-distance which is defined as:

$$r_e = \frac{d_e(\tilde{P}_k, [\underline{P}_j, \overline{P}_j])}{d_(P_j^n, \tilde{P}_k)}, \tag{10}$$

where d is the (precise) $\mathcal{L}_1$ distance between two patches and d_e is one of the extensions of the $\mathcal{L}_1$ distance ($e \in \{$Hausdorff, infinimum, mid, weighted $\}$. The distribution of the ratios for the different extensions are presented as a box-plot in Fig. 2. Table 1 shows the mean, median and standard deviation of the ratios distributions for each extension.

As might have been expected, the Hausdorff distance always leads to an over-evaluated distance, while, on contrary, the infimum distance leads to an under-evaluated distance. WVD and mid-point distance seem to provide distances that are consistent in that the means of their ratios are close to 1. For these extensions, comparing the imprecise patch with a precise one is statistically equivalent to comparing this precise patch with any patch contained in the imprecise patch. Although the ratios computed for the WVD have a slightly higher standard-deviation (which can be explained by the divergent behavior it has when some intervals tend to points) a Wilcoxon test applied on all the extensions showed the distribution of ratios of the WVD to be the closest to 1.

5 Conclusion

In this article we have introduced a new mathematical tool, the weighted variation of the mid-point distance, that allows to compare the performances of an interval-valued method with those of a conventional precise-valued method. We have presented its behavior through an experiment, where we compared it with other possible alternatives, namely the Hausdorff, mid-point and infimum distances. The WVD has some problems, such as its divergent behavior when some of the intervals radii tend to zero. However when considering a case of imprecise estimations where the imprecision has to be informative, meaning that the radius of the intervals should reflect the quality of the information provided, the new tool we proposed proved itself to have the best tradeoff between informativeness and consistency with the $\mathcal{L}_1$ distance.

Acknowledgements. This work is partially supported by TIN2014-56967-R (Spanish Ministry of Science and Innovation), FC-15-GRUPIN14-073 (Regional Ministry of the Principality of Asturias) and Merimee exchange program.

References

1. Beer, M., Ferson, S., Kreinovich, V.: Imprecise probabilities in engineering analyses. Mech. Syst. Signal Process. **37**, 4–29 (2013)
2. Aughenbaugh, J.M., Paredis, C.J.: The value of imprecise probabilities in engineering design. In: 2005 ASME Design Engineering Technical Conference and Computers and Information in Engineering Conference, American Society of Mechanical Engineers (ASME) (2005)
3. Oberguggenberger, M., King, J., Schmelzer, B.: Classical and imprecise probability methods for sensitivity analysis in engineering: a case study. Int. J. Approximate Reasoning **50**(4), 680–693 (2009). Imprecise Probability Models and their Applications (Issues in Imprecise Probability)
4. Afonso, M., Sanches, J.M.: Image reconstruction under multiplicative speckle noise using total variation. Neurocomputing **150**, 200–213 (2015)
5. de Souza, R.M., de Carvalho, F.: Clustering of interval data based on city-block distances. Pattern Recognit. Lett. **25**(3), 353–365 (2004)
6. Trutschnig, W., González-Rodrìguez, 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)
7. Augustin, T., Walter, G., Coolen, F.P.A.: Statistical inference, pp. 135–189 (2014)
8. Jaulin, L., Kieffer, M., Didrit, O., Walter, E.: Applied Interval Analysis. Springer, London (2001)
9. Mauris, G.: Possibility distributions: a unified representation of usual direct-probability-based parameter estimation methods. Int. J. Approximate Reasoning **52**, 1232–1242 (2011)
10. Strauss, O.: Non-additive interval-valued f-transform. Fuzzy Sets Syst. **270**, 1–24 (2015)