A Study of Distance/Similarity Measurements in the context of Signal Processing (Density Estimation)

I. A. Costa, D. M. Souza, R. A. Nóbrega

***Abstract*—Currently the number of applications where the data generation function is not known has been growing, making necessary the use of non-parametric estimation techniques to describe such model. Therefore, relevant questions emerge regarding the quality of the model that represents some dataset and how to quantify this quality. This article aims to evaluate some of the measurements presented in the literature used for this purpose, evaluating different pdf regions in the context of goodness of fit.**

***Index Terms*—Divergence, Similarity, Histogram, Probability Density Function.**

I. Introduction

In the last decades, several experiments generating large amounts of data were started, increasing the demand for studies related to stochastic probability density modeling using non-parametrics methods, where the probability density function (pdf) model is not known a priori. Therefore, to evaluate a pdf estimation procedure it becomes necessary goodness-of-fit measurements which may be done considering the distance or the similarity between functions.

Due to its importance, distance or similarity measurements between two functions are fundamental to improve the pdf estimation itself and to solve classification, clustering and spectral density estimation problems. However, there are a lot of distance and similarity measures in the literature [1] [2], each one more suitable for one kind of application than the other.

In [3], an annotated bibliography about divergence measurements for statistical data processing and inference problems was written. However, even after many studies, there is a continual demand to find an optimal measure of distance/similarity between functions for each kind of application and their particularities.

Motivated by the studies shown in [4], the main goal of this work is to find an efficient distance/similarity measurement to be applied to non-parametric methods of density estimation in the classification context. The second objective is to characterize the main distances/similarity measurements, assessing their sensitivities to separated pdf regions.

The paper is organized as follows. Section II presents the work methodology and its context. Section III shows the first approaches to a better understanding about the divergences and Section IV brings the achieved results and relevant comments on the subject. Finally, Section V concludes this work.

II. The Methodology

In the literature there are two ways to measure the difference of two objects, the distance and the similarity. The first one, satisfy the metric properties being called metric distance, and the other one, are non-metric distance, called divergence or similarity [4]. Distance and similarity measurements are denoted by d_x and s_x respectively throughout the paper.

The distance/similarity measures should be chosen considering which type of measurement will be analyzed. In this work it will be considered the most popular pattern representation, the probability density function in discrete form.

[4] has approached this subject analyzing various distance/similarity measures, dividing them into both syntactic and semantic relationships and measuring their correlation in order to know which distances/similarities might bring different information. By doing this, it is possible to select the most relevant distance/similarity measurements making it possible to perform a detailed study about them.

A. Distance/Similarity Measures

In order to study their characteristics, 42 measures have been selected as presented in Table I, following the same division proposed in [4]. Their equations can be found in [2].

Table I: Table of Divergence/Similarity divided by family

Divergence's Family			
L_p Minkowski	**L_1**	**Intersection**	**Inner Product**
L_1	Sorensen	Intersection	Inner Product
L_2	Gower	Wave	Harmonic
L_∞	Soergel	Motyka	Cosine
	Kulczynski	Kulczynski	Kumar
	Canberra	Ruzicka	Jaccard
	Lorentzian	Tanimoto	Dice
Fidelity or Squared-Chord	**Squared L_2**	**Shannon's Entropy**	**Combinations**
Fidelity	Squared Euclidean	Kullback-Leibler	Taneja
Bhattacharyya	Pearson χ^2	Jeffreys	Kumar-Johnson
Hellinger	Neyman χ^2	K Divergence	Avg(L_1,L_∞)
Squared-chord	Squared χ^2	Topsoe	
	ProbSym χ^2	Jensen-Shannon	
	Divergence	Jensen difference	
	Clark		
	Additive χ^2		

B. Dataset

We assume that our dataset, for the correlation test, is generated based on two distributions; a Gaussian $P = N(\mu, \sigma)$, where μ is mean or expectation of the distribution and σ is standard deviation, and a Gamma distribution $Q = P.\Gamma(\alpha, \beta)$, where α is a shape parameter and β is a rate parameter, denoted in this article by $\alpha = ne$ and $\beta = P_i/ne$. Figure 1 shows $P(0, 1)$ and Q for $ne = [50, 500, 5000, 50000]$ as representative example.

C. Regions of Interest

In order to understand the sensibility of each divergence/similarity measure in different regions of a pdf, three

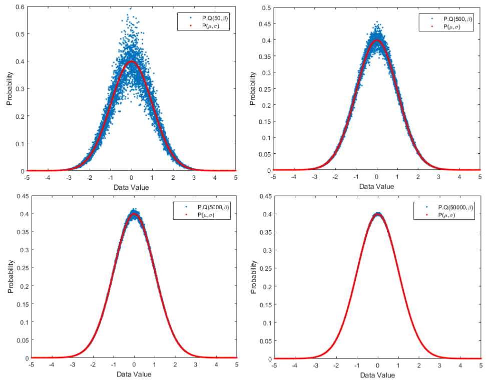


Figure 1: Four representative examples of $P(0,1)$ and Q for $ne = [50, 500, 5000, 50000]$, respectively.

regions of interest (RoI) are proposed: head, derivative and tail, as shown in Figure 2. Head has been defined as the region where the probabilities are greater than 0.9 of the maximum; tail less than 0.01 of the maximum and derivative as the region between head and tail. These RoI makes it possible to evaluate important specificities of the considered distance/similarity measures.

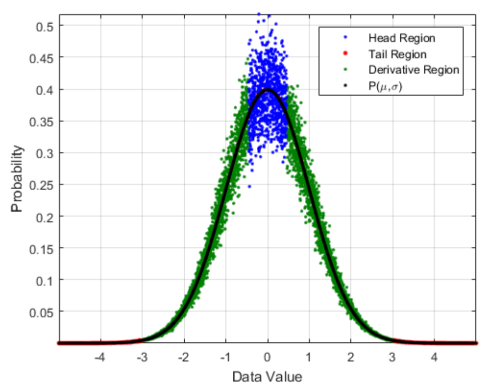


Figure 2: Representative figure of Regions of Interest.

D. Density Estimation

Density estimation can be done in different ways: Normalized Histogram, Normalized Average Shifted Histogram (ASH) and Kernel Density Estimation (KDE) are some of the main approaches in this matter.

1) Histogram: The estimation based on a histogram is one of the simplest ways to tackle this problem, naturally used in many fields, as represented by equation 1.

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(x_i \in B_j)\, I(x \in B_j) \qquad (1)$$

Where, n is the events number, i e j are the events subindices, I is the indicator function and,

$$\begin{aligned} &B_j = [x_o + (j-1)h, x_o + jh) \\ &j \in \mathrm{Z} \end{aligned} \qquad (2)$$

2) Average Shifted Histogram: The previous method handles with the origin parameter (x_o) and the bandwidth (h), both, if chosen wrongly, may result in poor estimation. In order to reduce this problem, estimation might be done by averaging sub-histograms, making the method independent of origin (x_o) [5], as shown in equation 3.

$$\hat{f}_h(x) = \frac{1}{M} \sum_{l=0}^{M-1} \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(x_i \in B_j)\, I(x \in B_j) \qquad (3)$$

Where,

$$\begin{aligned} &B_{j,l} = \left[\left(j - 1 + \tfrac{1}{M}\right)h, \left(j + \tfrac{1}{M}\right)h\right) \\ &l \in \{0, 1, ..., M-1\} \end{aligned} \qquad (4)$$

In this work, the bandwidth applied to the histogram and the ASH techniques was chosen based on the rule elaborated by Freedman-Diaconis [6].

3) Kernel Density Estimation: At last, a non-parametric technique known as Kernel Density Estimation [7], which general formulation is presented in equation 5, can be done using fixed or variable bandwidth, KDEs with variable bandwidth usually present better results in relation to those using fixed bandwidth [8] and can be found in detail here [9][10].

$$\hat{f}(x_i) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{(x_i - X_k)}{h}\right) \qquad (5)$$

where *K(u)* is the kernel function, *h* is the bandwidth and *n* is the number of events around the point x_i.

III. Divergence/Similarity Selection and RoI Evaluation Criteria

In this section it will be presented the procedure applied to select the divergences/similarities based on correlation, as described in [4], and the definition of a methodology used to compare the RoI in the context of divergence/similarity measurements.

A. Divergence/Similarity Selection

The correlation matrix considering all the 42 divergences is shown in Figure 3. The number of divergences can be reduced by choosing the measures with low correlation, keeping at least one of each family.

Table I could then be reduced to 12 divergences. The equation of each one of them is shown from Table II to Table VIII. Those divergences/similarities will be analyzed in details in the next sections.

Table II: L_p Minkowski family table

Divergence/Similarity	Equation	
L_∞	$d_{L_\infty} = \max_i \lvert P_i - Q_i \rvert$	(1)
L_{2Norm}	$d_{L_2} = \sqrt{\sum_{i=1}^{N} (P_i - Q_i)^2}$	(2)

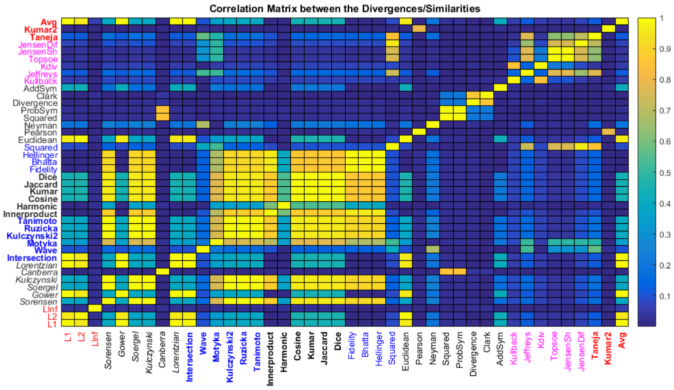


Figure 3: Correlation Matrix between the Divergences/Similarities.

Table III: L_1 family table

Divergence/Similarity	Equation	
$Sorensen$	$d_{sor} = \frac{\sum_{i=1}^{N} \lvert P_i - Q_i \rvert}{\sum_{i=1}^{N} (P_i + Q_i)}$	(3)
$Gower$	$d_{gow} = \frac{1}{N} \sum_{i=1}^{N} \lvert P_i - Q_i \rvert$	(4)

Table IV: Inner Product family table

Divergence/Similarity	Equation	
$InnerProduct$	$S_{IP} = \sum_{i=1}^{N} P_i Q_i$	(5)
$HarmonicMean$	$S_{HM} = 2 \sum_{i=1}^{N} \frac{P_i Q_i}{P_i + Q_i}$	(6)
$Cosine$	$S_{Cos} = \frac{\sum_{i=1}^{N} P_i Q_i}{\sqrt{\sum_{i=1}^{N} {P_i}^2} \sqrt{\sum_{i=1}^{N} {Q_i}^2}}$	(7)

Table V: Fidelity family table

Divergence/Similarity	Equation	
$Hellinger$	$d_H = \sqrt{2 \sum_{i=1}^{d} \left(\sqrt{P_i} - \sqrt{Q_i}\right)^2}$	(8)

Table VI: Squared L_2 family table

Divergence/Similarity	Equation	
$Squared\chi^2$	$d_{SqChi} = \sum_{i=1}^{N} \frac{(P_i - Q_i)^2}{P_i + Q_i}$	(9)
$AdditiveSymmetric\chi^2$	$d_{AdChi} = \sum_{i=1}^{N} \frac{(P_i - Q_i)^2 (P_i + Q_i)}{P_i Q_i}$	(10)

Table VII: Shannon's entropy family table

Divergence/Similarity	Equation	
$Kullback - Leibler$	$d_{KL} = \sum_{i=1}^{N} P_i \ln \frac{P_i}{Q_i}$	(11)

Table VIII: Combinations family table

Divergence/Similarity	Equation	
$Kumar - Johnson$	$d_{KJ} = \sum_{i=1}^{N} \left(\frac{({P_i}^2 - {Q_i}^2)^2}{2(P_i Q_i)^{3/2}} \right)$	(12)

B. RoI Evaluation Criteria

In order to create a controlled environment, a Monte Carlo simulation has been implemented: the estimation tests were carried out using a dataset generated from the analytical functions described in II-B. The dataset was generated based on a Gaussian function $N(1, 0)$ (to construct the pdf model) and an additional Gaussian noise. The analytical function used to generated the Gaussian pdf was then compared to its counterpart with additive noise and the RoI evaluated; each RoI has the same number of points.

Figures 4a and 4b show the normalized divergence measurements between these functions. Figure 4a presents the divergences obtained separately for each RoI, denoted as FIT, and Figure 4b presents the divergences obtained for the entire pdf applying the noise signal only to the RoI which is being evaluated, denote as $FULL$. Figures 4a and 4b show that only the divergences $Sorensen$, $InnerProduct$, $HarmonicMean$ and $Cosine$ present different results between these two methodologies.These differences can be explained by analyzing the equations of each divergence. The $Sorensen$ has a normalization dependence according to the number of points and the three others do not go to zero when one point is equal to the other, as shown in Figure 4c.

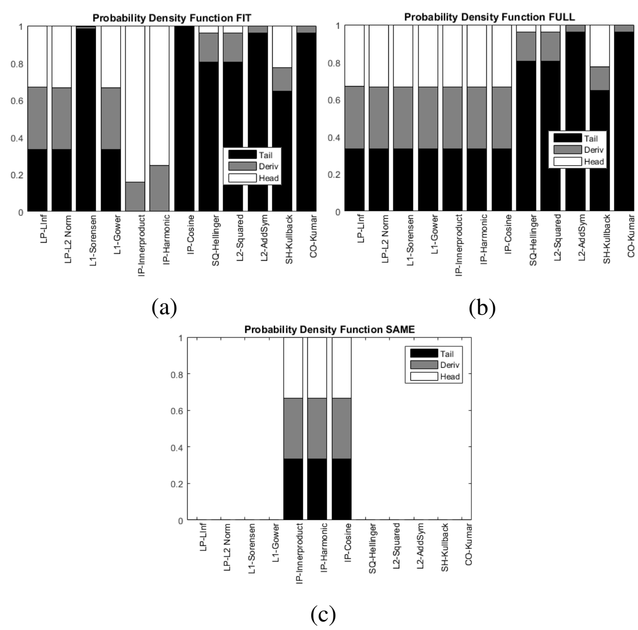


Figure 4: Difference between two functions. (a) Functions difference made separately for each region of interest, denote here as FIT, (b) Functions difference made for the entire PDF, but the uniform noise was applied in each region by time, denote here as $FULL$, and (c) The functions were the same.

Observing the absolute values resulted from applying these two methodologies to data and examining each divergence equation, the method which made it possible to normalize the results over the three RoI was chosen, according to Table IX.

Table IX: Methodology chosen for normalize each divergence/similarity

	Divergence/Similarity	Method
1	L_{Inf}	FULL
2	L_{2Norm}	*
3	$Sorensen$	FULL
4	$Gower$	FULL
5	$InnerProduct$	FIT
6	$HarmonicMean$	FIT
7	$Cosine$	FULL
8	$Hellinger$	*
9	$Squared\chi^2$	*
10	$AdditiveSymmetric\chi^2$	FIT/FULL
11	$Kullback-Leibler$	FIT/FULL
12	$Kumar-Johnson$	FIT/FULL

* They can not be normalized by any of the two methods because of the square root present in the equations.

IV. Results and Comments

Figure 5 shows the normalized divergences values achieved using the methodology defined in Table IX. One of the goals of this work is to clearly understand the sensitivity of distance measures according to the considered RoI. As it is possible to notice, divergences 1, 2, 3, 4 and 7 give equal importance to all RoI, whereas 5 and 6 are more sensitive to the head region, at the same time that 8, 9, 10, 11 and 12 prioritize the tail region.

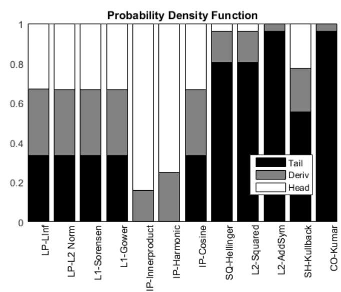
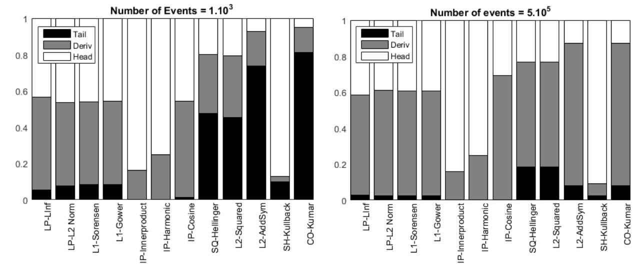


Figure 5: Difference between two functions.

To understand the effect of the binning process occurring in histograms construction, the Gaussian+noise pdf was binned based on [6], for a thousand and 500 thousands events. The results for this new dataset is shown in Figure 6. The results show that when binning is applied, the derivative region is significantly affected (the estimation error increases above the other RoI), followed by the head region.

To evaluate the effect of the number of events over the estimation process using the techniques described in section II (that might be understood as a Poisson process), this value has been scanned from 50 events to 100 thousands events. Figures 7 and 8 shows a representative case for the *Gower* divergence. It can be seen in Figures 7a and 7b that, in absolute values, the ASH presents lower values of divergence, as expected, and in Figures 7c and 7d it is possible to observe the convergence of the estimators in relation to the quantity of events. The Histogram tends to behave like the ASH as the number of events tends to infinity, whereas ASH with few events already has a reasonable response.



(a) (b)

Figure 6: Difference between two functions. (a) made with $1 * 10^3$ events number and (b) with $5 * 10^5$ events number.

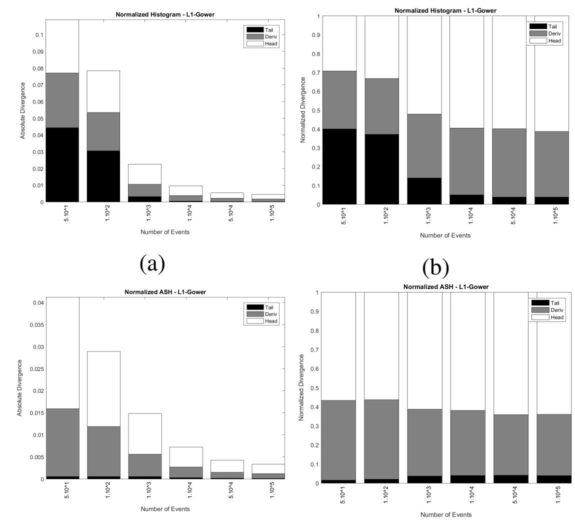
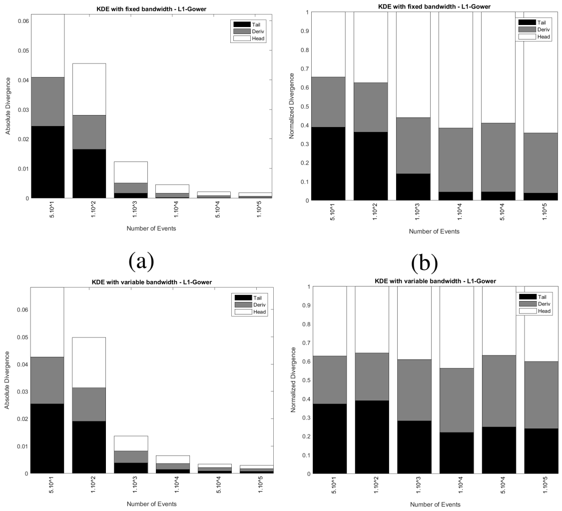


Figure 7: Divergence measurement of a density estimator for Gaussian distribution. (a) Absolute values for Normalized Histogram, (b) Normalized values for Normalized Histogram, (c) Absolute values for Normalized ASH and (d) Normalized values for Normalized ASH.

When Figures 8a and 8b are compared, it is possible to notice that, in absolute values, the two methods present similar results, for this reality, however, when looking at Figures 8c and 8d it is clear the effect of the variable bandwidth, which adjusts the bandwidth according to local probability event for each region keeping the contribution of each one similar to that shown in Figure 5.

Tables X, XI, XII and XIII show a summary of all the results considering all the divergences/similarities in the three regions of interest, in absolute and normalized values, for three different numbers of events and for the four considered estimation techniques: Histogram, ASH, KDE with fixed bandwidth and KDE with variable bandwidth, respectively. The measurements errors for the values shown in these tables are less than 3.5%.

Analyzing these results we can see that the divergences L_∞, L_2, Sorensen, Cosine, Hellinger and Squared χ^2 presents similar behavior to that described for the Gower divergence. The InnerProduct and Harmonic divergences present a certain



(a) (b)

(c) (d)

Figure 8: Divergence measurement of a density estimator for Gaussian distribution (a) Absolute values for KDE with fixed bandwidth, (b) Normalized values for KDE with fixed bandwidth, (c) Absolute values for KDE with variable bandwidth and (d) Normalized values for KDE with variable bandwidth.

immunity to the number of events and the histogram error, since the results shown in the Tables X, XI, XII and XIII resemble those of Figure 5. For the AddSym χ^2 divergence, the Histogram results did not converge to the ASH results and the KDE with fixed bandwidth seems to tend to give equal importance to the three RoI. Something interesting happened in one of the most used divergences in the literature, the Kullback-Leibler, besides the convergence of the Histogram towards the ASH results, and the absolute values are similar for both KDEs types, the fixed bandwidth KDE presented results that tends to the results found in the test shown in Figure 6; and the variable bandwidth KDE tends to equalize the three regions. Finally, the Kumar-Johnson similarity results show that the Histogram and the fixed bandwidth KDE tend to 33% in each region, however the ASH and the variable bandwidth KDE seem to give similar result to those found in the Gaussian noise analysis.

V. Conclusions

This work presented a careful evaluation of the main divergence/similarity measures present in the literature. Initially, 42 divergences were considered and the correlations between them were analyzed, reducing them to a total of 12. After that, their equations were revised, a method based on RoI was chosen and those measures were evaluated for a Gaussian density function. This study showed how each divergence behaves according to the different RoI. The effect of binning has also been analyzed in this same context.

The behavior of divergences/similarities in the density estimation process has also been evaluated using the concept of RoI. The achieved results provided a solid basis of knowledge

Table X: Table of normalized and absolute values of Histogram.

Divergence/ Similarity	Region of Interest	Number of Events $5*10^1$		$1*10^3$		$1*10^5$	
		ABS	NORM(%)	ABS	NORM(%)	ABS	NORM(%)
L_∞	Tail	0,191	29,32	0,015	7,64	0,002	3,69
	Deriv	0,275	44,59	0,104	53,03	0,027	49,46
	Head	0,162	26,09	0,078	39,33	0,026	46,85
L_2	Tail	92,743	38,21	6,655	12,51	0,414	3,74
	Deriv	76,978	33,23	19,442	37,65	4,196	37,62
	Head	66,696	28,56	26,505	49,84	6,637	58,64
Sorensen	Tail	0,092	37,97	0,007	13,81	0,000	3,80
	Deriv	0,073	31,01	0,017	33,83	0,004	34,86
	Head	0,076	31,02	0,028	52,36	0,007	61,34
Gower	Tail	0,044	39,93	0,003	13,95	0,000	3,80
	Deriv	0,033	30,84	0,007	33,90	0,002	34,86
	Head	0,032	29,24	0,012	52,15	0,003	61,34
Inner-Product	Tail	1,05E+02	0,10	8,45E+00	0,01	2,18E+00	0,00
	Deriv	1,88E+04	17,91	1,78E+04	16,28	1,77E+04	15,97
	Head	8,88E+04	81,99	9,15E+04	83,71	9,29E+04	84,03
Harmonic	Tail	1,51E+03	0,60	1,10E+03	0,43	7,62E+02	0,30
	Deriv	6,38E+04	25,60	6,22E+04	24,48	6,25E+04	24,41
	Head	1,85E+05	73,80	1,91E+05	75,09	1,93E+05	75,30
Cosine	Tail	0,043	46,08	0,000	5,93	0,000	0,31
	Deriv	0,028	34,81	0,002	37,91	0,000	30,24
	Head	0,015	19,12	0,003	56,17	0,000	69,45
Hellinger	Tail	282,973	56,41	59,413	46,73	6,973	32,86
	Deriv	150,428	29,62	36,769	30,60	7,445	35,33
	Head	70,351	13,97	27,356	22,67	6,758	31,81
Squared χ^2	Tail	220,059	51,97	51,737	43,98	6,785	32,31
	Deriv	136,078	31,81	35,679	31,82	7,441	35,61
	Head	69,971	16,23	27,337	24,20	6,758	32,08
AddSym χ^2	Tail	8,98E+06	96,07	5,65E+04	74,08	2,15E+02	34,11
	Deriv	2,99E+05	3,33	5,54E+03	16,11	2,24E+02	36,26
	Head	2,67E+04	0,59	3,26E+03	9,81	1,88E+02	29,63
Kullback-Leibler	Tail	2,94E+03	12,54	1,03E+03	16,33	4,76E+01	8,99
	Deriv	7,17E+03	18,87	1,08E+03	15,31	1,24E+02	12,65
	Head	3,03E+04	68,59	6,66E+03	68,36	7,08E+02	78,36
Kumar-Johnson	Tail	7,59E+07	98,86	1,48E+05	81,84	2,37E+02	36,04
	Deriv	7,39E+05	1,03	5,94E+03	11,49	2,24E+02	35,24
	Head	2,76E+04	0,12	3,27E+03	6,67	1,88E+02	28,72

Table XI: Table of normalized and absolute values of ASH.

Divergence/ Similarity	Region of Interest	Number of Events $5*10^1$		$1*10^3$		$1*10^5$	
		ABS	NORM(%)	ABS	NORM(%)	ABS	NORM(%)
L_∞	Tail	0,006	2,37	0,005	4,29	0,001	4,07
	Deriv	0,151	54,11	0,060	49,25	0,017	45,74
	Head	0,132	43,52	0,057	46,46	0,019	50,19
L_2	Tail	1,380	1,77	1,258	3,77	0,319	3,91
	Deriv	38,331	44,68	13,073	38,47	2,870	34,94
	Head	54,043	53,55	20,265	57,76	5,100	61,14
Sorensen	Tail	0,001	1,55	0,001	3,71	0,000	3,93
	Deriv	0,036	42,44	0,012	34,97	0,003	32,05
	Head	0,056	56,02	0,022	61,32	0,005	64,03
Gower	Tail	0,001	1,53	0,001	3,71	0,000	3,93
	Deriv	0,015	41,78	0,005	35,01	0,001	32,05
	Head	0,025	56,69	0,009	61,28	0,002	64,03
Inner-Product	Tail	1,12E+00	0,00	1,11E+00	0,00	2,18E+00	0,00
	Deriv	1,84E+04	15,72	1,77E+04	16,06	1,77E+04	15,97
	Head	1,01E+05	84,28	9,25E+04	83,94	9,29E+04	84,03
Harmonic	Tail	4,31E+02	0,17	3,17E+02	0,12	7,57E+02	0,30
	Deriv	5,76E+04	22,53	6,23E+04	24,47	6,25E+04	24,41
	Head	1,99E+05	77,31	1,92E+05	75,41	1,93E+05	75,29
Cosine	Tail	0,000	0,16	0,000	0,40	0,000	0,34
	Deriv	0,007	57,53	0,001	37,56	0,000	26,30
	Head	0,006	42,31	0,001	62,04	0,000	73,36
Hellinger	Tail	20,498	13,53	26,183	36,01	6,392	37,93
	Deriv	92,027	56,55	26,148	35,90	5,256	31,32
	Head	52,726	29,93	20,706	28,09	5,190	30,74
Squared χ^2	Tail	18,932	13,44	21,997	32,58	5,908	36,14
	Deriv	81,607	54,50	25,234	37,31	5,254	32,22
	Head	52,577	32,06	20,699	30,11	5,189	31,64
AddSym χ^2	Tail	3,49E+03	3,81	1,56E+04	62,66	9,49E+02	76,75
	Deriv	2,02E+05	86,68	1,00E+04	27,26	1,12E+02	11,81
	Head	1,58E+04	9,52	1,85E+03	10,08	1,11E+02	11,45
Kullback-Leibler	Tail	9,01E+02	3,47	1,59E+03	24,33	7,47E+01	13,03
	Deriv	1,72E+04	44,14	1,39E+03	17,21	1,29E+02	13,78
	Head	2,33E+04	52,39	6,11E+03	58,46	7,24E+02	73,19
Kumar-Johnson	Tail	5,49E+03	3,02	5,30E+04	71,46	4,85E+03	92,76
	Deriv	7,93E+05	92,36	4,75E+04	23,88	1,12E+02	3,71
	Head	1,61E+04	4,62	1,86E+03	4,66	1,11E+02	3,53

Table XII: Table of normalized and absolute values of KDE with fixed bandwidth.

Divergence/ Similarity	Region of Interest	Number of Events $5*10^1$ ABS	NORM(%)	$1*10^3$ ABS	NORM(%)	$1*10^5$ ABS	NORM(%)
L_∞	Tail	0,100	31,72	0,008	11,73	0,001	4,20
	Deriv	0,117	37,66	0,033	43,93	0,007	44,88
	Head	0,093	30,62	0,035	44,34	0,008	50,92
L_2	Tail	49,611	37,51	3,485	13,65	0,159	3,86
	Deriv	37,328	28,56	8,555	32,92	1,458	34,46
	Head	43,749	33,93	15,105	53,43	2,792	61,67
Sorensen	Tail	0,053	37,19	0,004	13,90	0,000	3,82
	Deriv	0,038	26,56	0,008	29,61	0,001	31,88
	Head	0,052	36,25	0,017	56,49	0,003	64,30
Gower	Tail	0,024	38,74	0,002	14,04	0,000	3,82
	Deriv	0,017	26,69	0,003	29,86	0,001	31,92
	Head	0,021	34,57	0,007	56,09	0,001	64,26
Inner-Product	Tail	5,84E+01	0,06	5,40E+00	0,01	2,27E+00	0,00
	Deriv	1,77E+04	17,49	1,77E+04	16,54	1,77E+04	16,04
	Head	8,43E+04	82,45	8,93E+04	83,46	9,25E+04	83,96
Harmonic	Tail	1,48E+03	0,60	1,07E+03	0,42	7,92E+02	0,31
	Deriv	6,39E+04	25,82	6,30E+04	24,91	6,26E+04	24,48
	Head	1,82E+05	73,58	1,89E+05	74,67	1,92E+05	75,21
Cosine	Tail	0,013	50,89	0,000	9,30	0,000	0,49
	Deriv	0,007	29,82	0,000	41,64	0,000	34,73
	Head	0,004	19,29	0,000	49,06	0,000	64,77
Hellinger	Tail	203,504	60,32	37,541	52,33	2,492	30,80
	Deriv	88,997	25,81	17,559	25,32	2,789	34,75
	Head	46,259	13,86	15,634	22,35	2,848	34,45
Squared χ^2	Tail	162,498	56,15	34,047	50,29	2,486	30,76
	Deriv	82,525	27,92	17,464	26,39	2,788	34,77
	Head	46,177	15,92	15,631	23,32	2,848	34,47
AddSym χ^2	Tail	2,46E+06	96,34	1,61E+04	77,74	2,74E+01	29,16
	Deriv	7,50E+04	2,93	1,38E+03	11,46	3,17E+01	34,98
	Head	1,05E+04	0,73	1,21E+03	10,80	3,63E+01	35,86
Kullback-Leibler	Tail	2,55E+03	10,07	6,99E+02	10,85	3,61E+01	5,19
	Deriv	4,32E+03	13,94	1,18E+03	12,70	1,88E+02	14,50
	Head	2,63E+04	75,99	8,24E+03	76,45	1,07E+03	80,31
Kumar-Johnson	Tail	1,44E+07	98,83	3,21E+04	83,01	2,76E+01	29,33
	Deriv	1,39E+05	0,98	1,41E+03	8,71	3,17E+01	34,89
	Head	1,06E+04	0,19	1,21E+03	8,29	3,63E+01	35,78

Table XIII: Table of normalized and absolute values of KDE with variable bandwidth.

Divergence/ Similarity	Region of Interest	Number of Events $5*10^1$ ABS	NORM(%)	$1*10^3$ ABS	NORM(%)	$1*10^5$ ABS	NORM(%)
L_∞	Tail	0,102	30,18	0,016	20,22	0,002	12,86
	Deriv	0,126	37,78	0,037	45,39	0,008	45,62
	Head	0,106	32,04	0,028	34,39	0,008	41,52
L_2	Tail	51,583	36,00	7,488	25,98	1,272	20,85
	Deriv	38,885	27,61	10,264	35,56	2,337	38,13
	Head	51,544	36,39	11,651	38,46	2,663	41,03
Sorensen	Tail	0,056	35,31	0,009	27,94	0,002	23,93
	Deriv	0,040	25,38	0,010	32,96	0,002	35,91
	Head	0,063	39,31	0,013	39,10	0,003	40,16
Gower	Tail	0,025	37,18	0,004	28,16	0,001	23,96
	Deriv	0,017	25,56	0,004	32,80	0,001	35,86
	Head	0,025	37,26	0,006	39,05	0,001	40,17
Inner-Product	Tail	6,10E+01	0,06	1,07E+01	0,01	3,81E+00	0,00
	Deriv	1,65E+04	17,09	1,69E+04	15,44	1,76E+04	15,94
	Head	8,10E+04	82,85	9,29E+04	84,55	9,30E+04	84,05
Harmonic	Tail	1,49E+03	0,62	1,29E+03	0,51	1,02E+03	0,40
	Deriv	6,21E+04	25,68	6,14E+04	24,06	6,22E+04	24,30
	Head	1,79E+05	73,70	1,93E+05	75,44	1,93E+05	75,30
Cosine	Tail	0,014	49,45	0,000	26,20	0,000	13,63
	Deriv	0,007	29,50	0,000	44,18	0,000	45,29
	Head	0,005	21,05	0,000	29,62	0,000	41,08
Hellinger	Tail	209,278	58,78	65,922	64,59	18,092	67,49
	Deriv	92,788	25,71	24,086	23,76	6,049	22,54
	Head	55,165	15,51	11,879	11,65	2,710	9,96
Squared χ^2	Tail	166,814	54,49	58,225	62,01	17,577	66,87
	Deriv	85,959	27,71	23,792	25,41	6,045	22,97
	Head	55,043	17,80	11,877	12,57	2,710	10,16
AddSym χ^2	Tail	2,61E+06	96,11	5,71E+04	93,35	1,59E+03	89,75
	Deriv	8,01E+04	3,02	2,60E+03	5,21	1,48E+02	8,35
	Head	1,44E+04	0,87	6,72E+02	1,43	3,37E+01	1,89
Kullback-Leibler	Tail	2,59E+03	8,76	1,32E+03	17,94	5,18E+02	34,77
	Deriv	2,81E+03	10,47	2,18E+03	28,30	5,25E+02	33,26
	Head	3,27E+04	80,76	5,10E+03	53,76	6,72E+02	31,97
Kumar-Johnson	Tail	1,54E+07	98,77	1,33E+05	96,36	1,81E+03	90,89
	Deriv	1,49E+05	1,01	2,77E+03	2,88	1,48E+02	7,43
	Head	1,46E+04	0,21	6,72E+02	0,76	3,37E+01	1,68

about these measures in many aspects, making this study a foundation for deeper investigations.

ACKNOWLEDGMENT

This work is supported by the Brazilian Ministry of Science, Technology, Innovation and Communication (MCTIC), FINEP, CNPq and FAPEMIG.

REFERENCES

[1] M.-M. Deza and E. Deza, *Dictionary of distances*. Elsevier, 2006.

[2] M. M. Deza and E. Deza, “Encyclopedia of distances,” in *Encyclopedia of Distances*. Springer, 2009, pp. 1–583.

[3] M. Basseville, “Divergence measures for statistical data processing—an annotated bibliography,” *Signal Processing*, vol. 93, no. 4, pp. 621–633, 2013.

[4] S.-H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *City*, vol. 1, no. 2, p. 1, 2007.

[5] S. Seather, “The performance of six popular bandwidth selection methods on some real data sets,” *COMPUTATIONAL STATISTICS QUARTERLY*, vol. 7, pp. 225–225, 1992.

[6] D. Freedman and P. Diaconis, “On the histogram as a density estimator: L 2 theory,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, 1981.

[7] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[8] G. R. Terrell and D. W. Scott, “Variable kernel density estimation,” *The Annals of Statistics*, pp. 1236–1265, 1992.

[9] I. A. Costa, D. M. Souza, R. A. Nóbrega, and J. M. Seixas, “Implementação e avaliação de um algoritmo de estimação de densidades baseado no mkde aplicado à classificação de elétrons no detector atlas,” *XXI Congresso Brasileiro de Automática*, pp. 2902–2907, 2016.

[10] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.