



## A method for continuous-range sequence analysis with Jensen-Shannon divergence

M. A. Ré<sup>1,2\*</sup>, G. G. Aguirre Varela<sup>2,3†</sup>

Mutual Information (MI) is a useful Information Theory tool for the recognition of mutual dependence between data sets. Several methods have been developed for estimation of MI when both data sets are of the discrete type or when both are of the continuous type. However, MI estimation between a discrete range data set and a continuous range data set has not received so much attention. We therefore present here a method for the estimation of MI for this case, based on the kernel density approximation. This calculation may be of interest in diverse contexts. Since MI is closely related to the Jensen Shannon divergence, the method developed here is of particular interest in the problems of sequence segmentation and set comparisons.

### I Introduction

Mutual Information (MI) is a quantity whose theoretical base originates in Information Theory[1]. Since MI between two independent random variables (RV) is zero, a non-null value of MI between these variables gives a measure of mutual dependence. When analyzing two data sets  $X$  and  $Y$  (assumed to be the realization of two mutually dependent RVs) MI can give us a measure of the

mutual dependence of these sets. Although MI may be straightforwardly calculated when the underlying probability distributions are known, this is not usually the case when only the data sets are available. Therefore, MI must be estimated from the data sets themselves. When  $X$  and  $Y$  are the discrete type, MI may be estimated by substituting the joint probability of these variables by the relative frequency of appearance of each pair  $(x, y)$  in the data sequence [2, 3]. For real value data sets (or the discrete type with a wide range) estimation of MI by frequency of appearance is not applicable. The binning method [4] in turn requires large bins or large sequences in order to produce reasonable results. Alternative proposals have been made for cases when both data sets are the continuous type [5].

<sup>1</sup> Centro de Investigación en Informática para la Ingeniería, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Maestro López esq. Cruz Roja Argentina, (5016) Córdoba, Argentina.

<sup>2</sup> GFA - Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Av. Medina Allende s/n, Ciudad Universitaria, (5000) Córdoba, Argentina.

<sup>3</sup> Instituto de Física Enrique Gaviola (IFEG), Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Ciudad Universitaria, (5000) Córdoba, Argentina.

Estimation of MI between a discrete RV and a continuous one has not been so extensively considered, in spite of being a problem of interest in diverse situations. For instance, we could compare the day of the week (weekday-weekend, discrete) with traffic flow (continuous), quantifying this effect. In a different context we might wish to quan-

tify the effect of a drug (administered or not, discrete) in medical treatment evaluation (electroencephalograms in epilepsy, continuous data). Ross[6] has proposed a scheme for estimating MI based on the nearest neighbour method [4]. Assuming a sequence of  $(x, y)$  pairs, with  $X$  being discrete and  $Y$  continuous, the nearest neighbour method requires the pairs to be ordered by the  $Y$  values. This requirement makes the proposal impractical, in sequence analysis for instance. The nearest neighbour method was also considered by Kraskov *et al.* [7]. In their paper they suggest two ways of evaluating MI with this method. An alternative definition for MI is presented by Gao *et al.* [8], also based on the distance between the elements of the sequence.

In this paper we propose a more direct method for estimating MI between a discrete and a continuous data set, based on the kernel density approximation (KDA)[4] for estimating the probability density function (PDF) of the continuous variable. For the discrete variable we make use of the usual frequency approximation [2,3]. Finally, MI is computed by the Monte Carlo integration.

As shown by Grosse *et al.* [2] MI can be identified with the Jensen Shannon Divergence (JSD), a measure of dissimilarity between two probability distributions. JSD is a non-negative functional that equals zero when the distributions being compared are the same. This property makes JSD a useful tool for sequence segmentation [2,3]. Furthermore, in diverse contexts it is of interest to evaluate whether a given sequence matches a particular probability distribution. The most usual case is that of a normal distribution. Nevertheless, this is a more general problem. For instance, in satellite synthetic aperture radar (SAR) images the backscatter presents a multiplicative noise assumed to have an exponential distribution [9]. Also, models for cloud droplet spectra assume a Weibull distribution [10,11]. Several indirect methods have been proposed for analysis of continuous range sequences. Pereyra *et al.* [12] outlined a method based on wavelet transform to analyze electroencephalograms. Recently, Mateos *et al.* [13] have proposed a mapping from continuous value sequences into discrete state sequences previous to JSD calculation. Several other mapping methods have been proposed in the literature to associate a discrete probability distribution with a real value series.

Here, by means of the KDA we avoid resorting to any indirect method, approximating the probability densities of continuous range variables by this non parametric method. In section II we present the calculation of MI and the arrangement for sequence segmentation with JSD. In section III we test the performance of this method through numerical experiments. Also considered is application of the method in edge detection in a satellite synthetic aperture radar (SAR) image. In section IV we consider the results obtained.

## II Method

In this section we present our proposal for estimating MI between discrete and continuous RVs, based on the KDA estimator of a PDF. Let us consider a sequence of pairs  $(x, y)$  with  $x$  as a variable of discrete range and  $y$  of continuous range. To calculate MI we resort only to the sequence itself, making use of no extra information. We start from the sequence of data pairs  $(x, y)$ , and assume that these data are sampled from a joint probability density  $\mu(x, y)$ , although unknown at first. From the joint PDF the marginal probabilities

$$p(x) = \int_{-\infty}^{\infty} dy \mu(x, y) \quad (1a)$$

$$\phi(y) = \sum_x \mu(x, y) \quad (1b)$$

are defined.

The MI between the RVs  $X$  and  $Y$  is expressed in terms of these PDFs as [1]

$$I(X, Y) = \sum_x \int_{-\infty}^{\infty} dy \mu(x, y) \ln \left[ \frac{\mu(x, y)}{p(x) \phi(y)} \right]. \quad (2)$$

Note that if the variables  $X$  and  $Y$  are statistically independent then  $\mu(x, y) = p(x) \phi(y)$ , and in this case  $I(X, Y) = 0$ . In this way a value of  $I(X, Y) \neq 0$  gives a measure of the mutual dependence of these variables. We may rewrite  $I(X, Y)$  in terms of the conditional PDFs

$$\mu(y | x) = \frac{\mu(x, y)}{p(x)} \quad (3)$$

as

$$I(X, Y) = \sum_x p(x) \int_{-\infty}^{\infty} dy \mu(y | x) \ln \left[ \frac{\mu(y | x)}{\phi(y)} \right]. \quad (4)$$

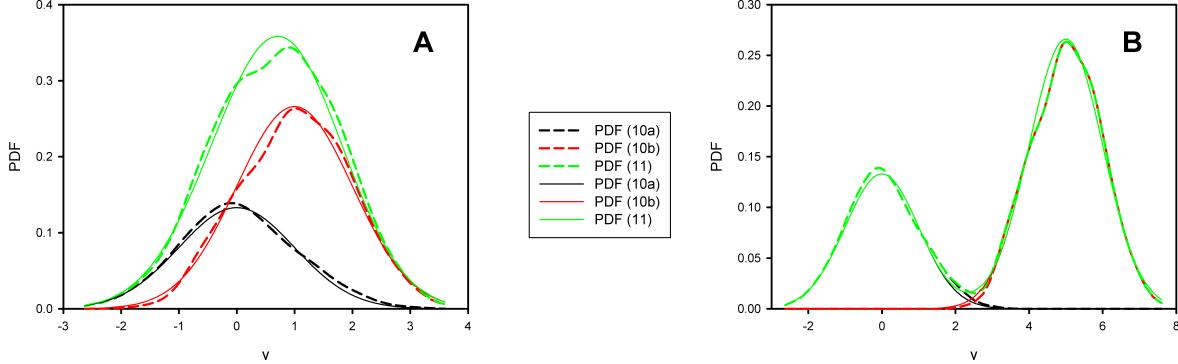


Figure 1: Kernel Density Approximation (KDA) for the Probability Density in (10) calculated from 1000 pairs generated by the Monte Carlo method. For plot A  $y_m = 1$ , while for plot B  $y_m = 5$ . In both cases  $\sigma_g = 1$ . Solid lines correspond to the analytic function and dashed lines to the KDA.

### i Kernel density approximation

To carry out the calculation in (4), knowledge of the conditional PDFs is necessary. As mentioned, these densities are assumed to be unknown, and have to be estimated from the data themselves. Here we make use of the KDA [4], as summarized in the following. The conditional PDFs in Eq. (3) are estimated considering separately each data set pair with a given value of  $x$ . We define the set

$$\mathcal{C}_\kappa = \{(x, y) / x = \kappa\} \quad (5)$$

and for each set we approximate the conditional densities using a KDA with a Gaussian kernel

$$\hat{\mu}(y | x = \kappa) = \frac{1}{n_\kappa h_\kappa} \frac{1}{\sqrt{2\pi}} \sum_{y_j \in \mathcal{C}_\kappa} \exp \left[ -\frac{(y - y_j)^2}{2h_\kappa^2} \right]. \quad (6)$$

Note that the sum is over the  $y_j$  values in the set  $\mathcal{C}_\kappa$ , and  $n_\kappa$  is the number of pairs in this set. The bandwidth,  $h_\kappa$ , is chosen as the optimal value, as reported by [4] and followed by Steuer *et al.* [5]

$$h_\kappa \simeq 1.06 s_\kappa n_\kappa^{-0.2} \quad (7)$$

where  $s_\kappa^2$  is the variance of the sample. Sheather [14] considered alternative values to detect bimodality; however, as they mention, there is little visual difference.

The marginal probability of  $X$  is approximated by the frequency of occurrence of each value

$$\hat{p}(x = \kappa) = \frac{n_\kappa}{n} \quad (8)$$

and the marginal probability density of  $Y$  by

$$\hat{\phi}(y) = \sum_x \hat{p}(x) \hat{\mu}(y | x). \quad (9)$$

We illustrate the results obtained with the KDA by an example: let us consider the joint probability distribution  $\mu(x, y)$

$$\mu(x = 1, y) = \frac{1}{3} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{y^2}{2} \right] \quad (10a)$$

$$\mu(x = 2, y) = \frac{2}{3} \frac{1}{\sqrt{2\pi}\sigma_g} \exp \left[ -\frac{(y - y_m)^2}{2\sigma_g^2} \right] \quad (10b)$$

and the corresponding marginal PDF

$$\phi(y) = \frac{1}{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} + \frac{2}{3} \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\frac{(y - y_m)^2}{2\sigma_g^2}}. \quad (11)$$

We sampled 1000 pairs from this distribution for two different values of  $y_m$ , and from these pairs we made an estimation of the conditional PDFs using the KDA. In Fig. 1A and 1B we plot the probability functions in (10) and (11) for two values of  $y_m$  and the corresponding approximations.

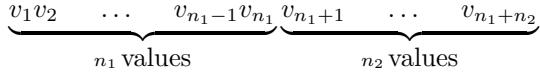


Figure 2: *The segmentation problem.* Consider a sequence  $\mathcal{S}$  made up of two stationary subsequences  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , with  $n_1$  and  $n_2$  elements respectively. The problem consists in determining the value of  $n_1$ ; *i.e.*, the point when the statistical properties change.

## ii Monte Carlo integration

After approximating the PDFs we have to compute the integrals in (4) to estimate MI. We recognize in these integrals the expectation value

$$\left\langle \ln \frac{\mu(y|x)}{\phi(y)} \right\rangle = \int_{-\infty}^{\infty} dy \mu(y|x) \ln \left[ \frac{\mu(y|x)}{\phi(y)} \right] \quad (12)$$

that can be estimated by Monte Carlo integration [15]

$$\left\langle \ln \frac{\mu(y|\kappa)}{\phi(y)} \right\rangle \simeq \frac{1}{n_\kappa} \sum_{y_j \in \mathcal{C}_\kappa} \ln \left[ \frac{\hat{\mu}(y_j|x=\kappa)}{\hat{\phi}(y_j)} \right]. \quad (13)$$

Here the sum is again restricted to the  $y_j$  values associated with a particular  $x$  value. Note that in this sum we make use of the kernel approximation of the conditional PDFs in (6). Substituting both approximations we finally get

$$\hat{I}(X, Y) \simeq \frac{1}{n} \sum_x \sum_{y_j \in \mathcal{C}_\kappa} \ln \left[ \frac{\hat{\mu}(y_j|x)}{\hat{\phi}(y_j)} \right]. \quad (14)$$

## iii Sequence segmentation

The JSD is a measure of dissimilarity between probability distributions. Originally proposed by Burbea and Rao [16] and Lin [17] as a symmetrized version of Kulback Leibler divergence [1, 18], a generalized weighted JSD between two PDFs,  $f_1, f_2$  is defined as

$$D[f_1, f_2] = H(\pi_1 f_1 + \pi_2 f_2) - \pi_1 H(f_1) - \pi_2 H(f_2) \quad (15)$$

with  $\pi_i$  arbitrary weights satisfying  $\pi_1 + \pi_2 = 1$ . Here  $H$  is Gibbs Shannon entropy, defined for continuous range variables as

$$H(f_i) = - \int_{-\infty}^{\infty} dy f_i(y) \ln [f_i(y)]. \quad (16)$$

As shown by Grosse *et al.* [2] JSD may be interpreted as MI between a discrete and a continuous variable by identifying the weights  $\pi_i$  with the discrete variable probability in (1a):

$$\pi_i = p(x=i) \quad (17)$$

and the probability densities  $f_i(y)$  with the conditional densities in (3)

$$f_i(y) = \mu(y|x=i). \quad (18)$$

With these identifications, the functionals in (15) and (4) are the same.

The JSD and several generalizations have been successfully applied to the sequence segmentation problem, the partition of a non-stationary sequence into stationary subsequences, for discrete range sequences. We propose here the extension of this method to continuous range sequences without resorting to discrete mapping, wavelet decomposition or any other indirect method of estimation of the probability distribution.

The procedure for sequence segmentation may be stated in the following way: let us consider a sequence  $\mathcal{S}$  with  $n$  elements made of two stationary subsequences  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , with  $n_1$  and  $n_2$  values respectively ( $n_1 + n_2 = n$ ), schematically illustrated

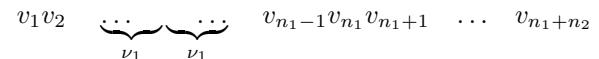


Figure 3: *The sliding window method.* A sliding window is defined for sequence segmentation. The window is divided into two subwindows of equal size. The center of the window is considered as the window position. The window is displaced along the sequence and the JSD between the subwindows is calculated. The segmentation point is identified as the window position at which JSD has its maximum value.

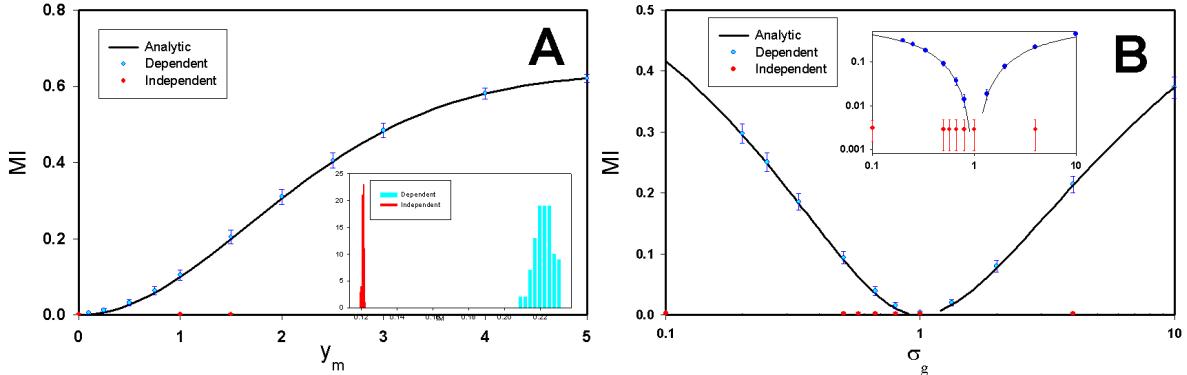


Figure 4: Mutual information estimation for the joint distribution in (10). For the distribution in (10) the dots represent the average MI value for 100 data sets of 1000 ( $x, y$ ) pairs each, with the bars indicating the standard deviation of each set. The black line is the analytical value of MI: A) as a function of the mean value  $y_m$  in (10b)(the inset shows the distribution of MI for a particular value of  $y_m$  for a dependent and an independent set), and B) changing  $\sigma_g$ , the standard deviation in (10b). The inset shows the same plot but in log-log scale to highlight the MI value for independent sets.

in Fig. 2. The aim is to determine the value of  $n_1$ ; *i.e.*, the position of the last element in  $\mathcal{S}_1$ . In the algorithm proposed here we define a sliding window of fixed width over the sequence. The window is divided into two segments, each including  $\nu_1$  elements (see Fig. 3). We define the window position as that of the last element in the left section of the window. This window is displaced over the sequence and the window position where JSD reaches its maximum value is taken as the segmentation point.

### III Assessment results

In this section we present the results of our assessment of the proposed method by considering two applications: the detection of mutual dependence between two RV sequences and the segmentation of a sequence.

In the first case we generate sequences of two jointly distributed variables: one of discrete range and one of continuous range, and then we compute MI between these variables. In the second case we consider sequences made of two subsequences generated from different distributions. We detect the segmentation point following the procedure described in the previous section. We also apply the method to detect the edges between homogeneous regions in SAR images.

#### i Mutual information between a discrete and a continuous variable

We computed the MI between discrete and continuous variables. We generated 100 data sets, sampling 1000 ( $x, y$ ) pairs from the distribution in (10) with different values of  $y_m$  or  $\sigma_g$ , and from the joint distribution

$$\mu(x=1, y) = \frac{1}{3} [\Theta(y-0.5) - \Theta(y+0.5)] \quad (19a)$$

$$\mu(x=2, y) = \frac{2}{3} \frac{1}{a} \left[ \Theta\left(y + y_m - \frac{a}{2}\right) - \Theta\left(y + y_m + \frac{a}{2}\right) \right] \quad (19b)$$

with  $\Theta(y)$  the step function

$$\Theta(y) = \begin{cases} 0 & \text{for } y < 0 \\ 1 & \text{for } y > 0 \end{cases} \quad (20)$$

with different values of  $y_m$  or  $a$ . We estimated the MI,  $I(X, Y)$ , from each set by the method described in the previous section. Given that we are sampling the data pairs from known distributions, we are also able to calculate MI from the analytical expressions. In this way we may compare the results obtained from the approximation with the corresponding analytical results.

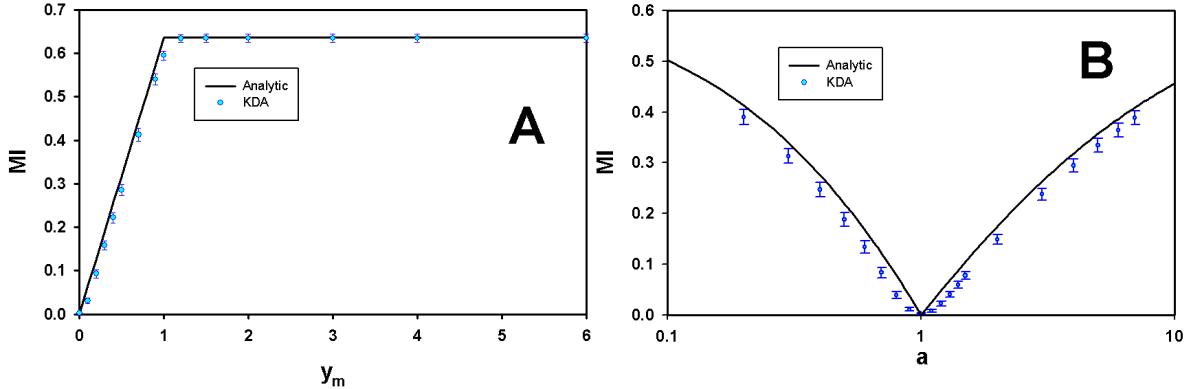


Figure 5: Mutual information estimation for the joint distribution in (19). For the distribution in (19) the dots represent the average MI value for 100 data sets of 1000 ( $x, y$ ) pairs each, with the bars indicating the standard deviation. The black line is the analytical value of MI while the dots represent the Kernel Density Approximation (KDA) values; A) as a function of mean value  $y_m$  in (19b), and B) changing the width parameter  $a$  in (19b).

In addition, we calculated the MI for samples of statistically independent variables to establish a significance value for the MI of the dependent variables. The analytical value in this case is zero, as already mentioned. The results of the calculation are shown in Fig. 4 for the distribution in (10) and in Fig. 5 for the distribution in (19), respectively. We include the average value of MI over the 100 data sets for the different values of the parameters, and the bars correspond to the standard deviation in each set. A small underestimation of the MI value can be seen in this last case. This may be attributed to a shortcoming of the KDA at the borders of the interval of the uniform distribution. Nevertheless, it is still possible to detect mutual dependence between the discrete and the continuous value sequences.

To consider the effect of sample size, we repeated the experiment with the distribution in (10) for different values of  $n$ , the number of data pairs in each set. We again generated 100 data sets of  $n$  data pairs each. The results are shown in Fig 6 for three sets of parameters. A slightly increasing overestimation of MI can be appreciated as  $n$  decreases. Finally, we considered an usual situation when there is only one sample of data pairs available. We sampled 1000 pairs from the distribution in (10), the distribution in (19) and from the distribution

$$\begin{aligned} \mu(x=1, y) &= \frac{1}{3} \exp(-y) \Theta(y) \\ \mu(x=2, y) &= \frac{2}{3} \exp(-y/2) \Theta(y). \end{aligned} \quad (21)$$

For each sample we estimated MI by the approximate method in (14). To set up a significance value

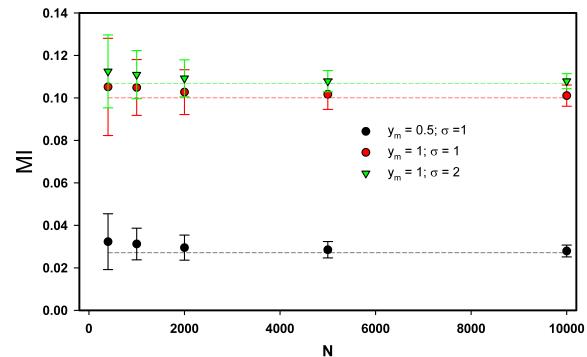


Figure 6: Mutual information estimation for the distribution in (10). For the distribution in (10) the dots represent the average value for 100 data sets of different numbers of ( $x, y$ ) pairs. Bars indicate the standard deviation, and dashed lines represent the analytical values of MI for the different sets of parameters.

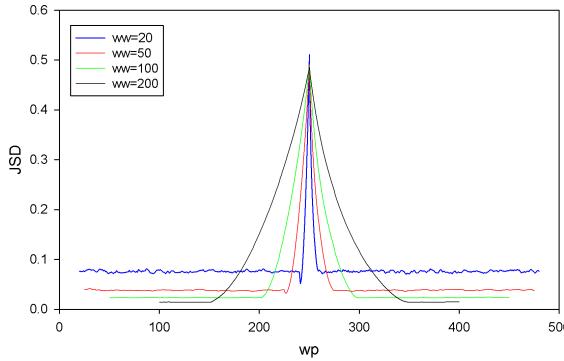


Figure 7: *Segmentation point in artificial sequences.* The JSD average computed for 500 sequences generated from Rayleigh distributions. Each sequence has a length of 500 elements divided into two subsequences with 250 elements each. The ratio of the mean values of the subsequences is given by  $r_m = m_l/m_r = 5$ , where  $m_l$  and  $m_r$  are the mean values in the left and right subsequences, respectively. The sequences are analyzed with different window widths (ww). In all cases the window position (wp) of the maximum JSD average is at the segmentation point.

for each sample we generated 100 data sets of 1000 pairs of independent variables. The discrete values were sampled from the distribution

$$p(x) = \frac{n_x}{1000} \quad (22)$$

where  $n_x$  gives the number of times that the value  $x$  appears in the original sequence, and the continuous values were sampled from the Gaussian distribution

$$\mu(y) = \frac{1}{\sqrt{2\pi}s} \exp\left[-\frac{(y-m)^2}{2s^2}\right] \quad (23)$$

independently of the value of  $x$ . Here  $m$  is the mean value in the original sequence and  $s^2$  the sample variance. We calculated the MI for each data set and then the MI mean value and its variance. The results are included in Table 1. A clear difference can be seen between the MI of the dependent values and those of the independent sequences.

## ii Sequence segmentation

To test the sequence segmentation method, we generated sets of 500 sequences of 500 values each, di-

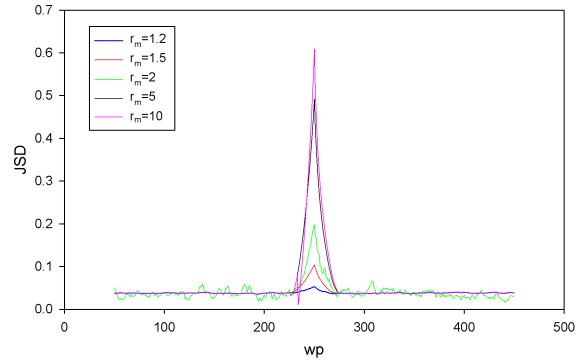


Figure 8: *Segmentation point in artificial sequences.* The JSD average computed for 200 sequences generated from Rayleigh distributions. Each sequence has a length of 500 elements divided into two subsequences with 250 elements each. Different values of the mean quotient  $r_m = m_r/m_l$  are considered, where  $m_r$  is the mean value of the right subsequence and  $m_l$  the mean value of the left subsequence. In all cases a window width of 50 elements was used. Even for the lowest quotient value the window position (wp) of the maximum JSD average is coincident with the segmentation point.

vided into two subsequences with 250 values in each one. The sequences were generated from Rayleigh distributions with a different mean value for each subsequence. The mean value of the first subsequence is denoted by  $m_l$ , and the mean value of the second segment by  $m_r$ ; we define the ratio of the mean values as  $r_m = m_l/m_r$ . Using the sliding window method, we analyzed a set with  $r_m = 5$  with several window widths. In Fig. 7 we present the average value across the 500 sequences of JSD at each window position for the different widths considered. The average JSD has

Table 1: *Mutual information and significance value.* MI of the sampled dependent sequences (see text) and the corresponding significance values computed from the independent sets.

PDF	MI	Significance Value	
		mean	st. dev.
Gaussian	0.6359	$4.5 \times 10^{-3}$	$1.8 \times 10^{-3}$
Uniform	0.1429	$4.5 \times 10^{-3}$	$1.8 \times 10^{-3}$
Exponential	0.0718	$4.5 \times 10^{-3}$	$1.9 \times 10^{-3}$

a maximum value at position 250, the segmentation point, even for a narrow window with 20 elements (10 elements in each subwindow), although in this case statistical fluctuations are more noticeable. To test the sensitivity of the method we generated sets with  $r_m = 1.2, 1.5, 2, 5, 10$ . The results of the algorithm, with a window of 50 elements, are included in Fig. 8. Even for the smallest ratio considered, the segmentation point can be detected.

Finally, we present an example of application of the segmentation algorithm to detect the edge between homogeneous regions in a SAR image. In SAR images the backscatter is affected by speckle noise (a multiplicative noise). This noise in the backscatter amplitude is modelled by a Rayleigh distribution in homogeneous regions. In Fig. 9 we include a section of the SAR image of an Antarctic region, and the boundary detected between water and ice. On the right a plot of the values of the backscatter amplitude of the highlighted lines in the image and the JSD is included. There is good coincidence of the detected boundary with the contour in the image.

#### IV Discussion and conclusions

In this paper we have presented a method for computing Mutual Information (MI) between discrete and continuous data sets, or alternatively, the JSD between continuous range data sets. The algorithm developed gives a measure of dissimilarity without resorting to an indirect method like those proposed in [12, 13]. Neither is it necessary to have the continuous values ordered as in the nearest neighbour method [4, 6]. In fact, the calculation in (14) is based only on the registered data as they were recorded.

The measure may be applied to two similar problems. On the one hand we can quantify the mutual dependence between discrete and continuous data sets, and on the other hand we can quantify the dissimilarity between two continuous data sets, as discussed in section II. In section III we applied the method to artificially-generated pairs of variables, finding good agreement with the corresponding analytical values as shown in Figs. 4 and 5, although systematic underestimation occurs mainly when the difference is given by the width in uniform distributions (fig 5-B). We attribute this dis-

crepancy to the abrupt decay of the uniform distribution at the borders of the interval, while the KDA with a Gaussian kernel extends to infinity. The MI values in these cases of mutually dependent variables are clearly distinguishable from the MI values of independent variables. We also considered the dependence of the results of this method on the length of the sequence. In Fig 6 a slightly increasing overestimation of MI is seen with decreasing length. Nevertheless, there is good agreement for sequences of more than 400 pairs.

In real situations we frequently have only one sequence of  $(X, Y)$  pairs. We have proposed a method for establishing a significance value by generating 100 sequences of independent variables with probability distributions given by the estimated marginal distribution for the discrete variable, and by a Gaussian distribution for the continuous variable with the same mean value and variance as the marginal distribution of the original sequence. We have considered sequences generated from three distributions. In all three cases MI establishes a clear difference between dependent and independent sets, as shown in Table 1.

It has been shown that the Jensen Shannon Divergence (JSD) is equivalent to MI [2]. Therefore, the calculation method developed here will also be suitable for computing JSD between two continuous range data sets, and in this format the JSD may be applied to the sequence segmentation problem as proposed in section II-iii. In this section we suggested a method based on a fixed-length sliding window. We considered the segmentation of artificially generated sequences in section III-ii. The JSD average at each position in the sequences exhibits a maximum at the segmentation point, as shown in Fig. 7. As we continue this work we will address the problem of comparing and analyzing electrophysiological signals. The segmentation method may also be of interest in detecting borders in images. Work along these lines will be published elsewhere.

*Acknowledgements* - We wish to acknowledge partial support from SCyT - UTN through grant UTI4811 and from SeCyT - UNC through grant 30720150100199CB.

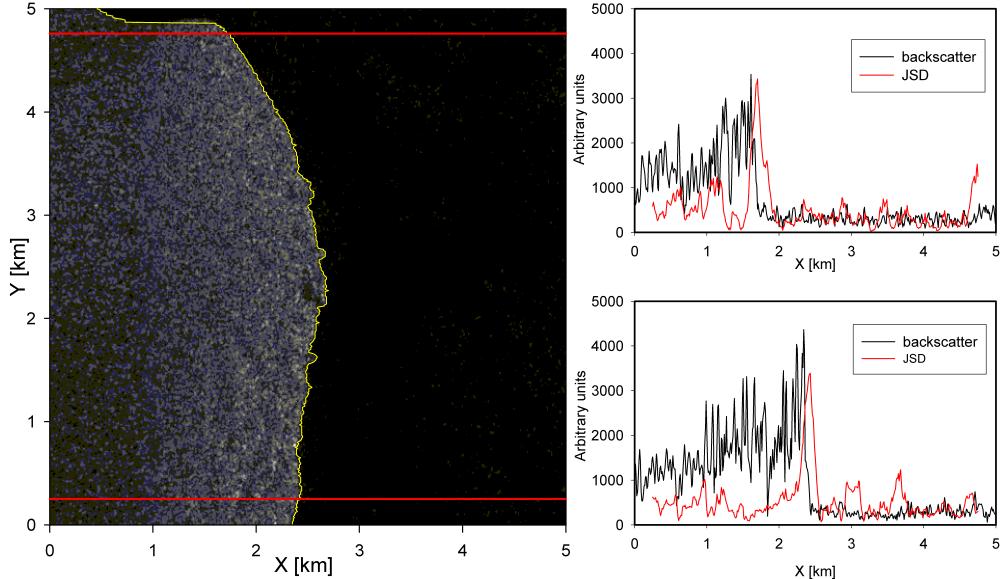


Figure 9: *Border detection in SAR images.* The segmentation method was applied to detection of the border between homogeneous regions in a SAR image. The image was analyzed line by line and the segmentation point at each line detected. The segmentation points are coincident with the border.

- 
- [1] T Cover, J Thomas, *Elements of Information Theory*, J. Wiley, New York (2006).
  - [2] I Grosse, P Bernaola-Galván, P Carpena, R Román-Roldán, J Oliver, H. E. Stanley, *Analysis of symbolic sequences using the Jensen-Shannon divergence*, Phys. Rev. E, **65**, 041905 (2002).
  - [3] M A Ré, R K Azad, *Generalization of entropy based divergence measures for symbolic sequence analysis*, PLoS ONE **9**, e93532 (2014).
  - [4] B W Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, London (1986).
  - [5] R Steuer, J Kurths, C O Daub, J Weise, J Selbig, *The mutual information: Detecting and evaluating dependencies between variables*, Bioinformatics **18**, S231 (2002).
  - [6] B C Ross, *Mutual Information between discrete and continuous data sets*, PLoS ONE **9**, e87357 (2014).
  - [7] A Kraskov, H Stögbauer, P Grassberger, *Estimating mutual information*, Phys. Rev. E **69**, 066138 (2004).
  - [8] W Gao, S Kannan, S Oh, P Viswanath *Estimating mutual information for discrete-continuous mixtures*, 31<sup>st</sup> Conference on neural information processing systems (NIPS), 5986 (2017).
  - [9] A Moreira, P Prats-Iraola, M Younis, G Krieger, I Hajnsek, K P Papathanassiou, *A tutorial on Synthetic Aperture Radar*, IEEE Geosci. Remote S. Magazine **1**, 6 (2013).
  - [10] Y Liu, J Hallett, *On size distributions of cloud droplets growing by condensation: a new conceptual model*, J. Atmos. Sci. **55**, 527 (1998).
  - [11] Y Liu, P H Daum, J Hallett, *A generalized systems theory for the effect of varying fluctuations on cloud droplet size distributions* J. Atmos. Sci. **59**, 2279 (2002).

- [12] M E Pereyra, P W Lamberti, O A Rosso, *Wavelet Jensen-Shannon divergence as a tool for studying the dynamics of frequency band components in EEG epileptic seizures*, *Phys. A* **379**, 122 (2007).
- [13] D M Mateos, L E Riveaud, P W Lamberti, *Detecting dynamical changes in time series by using Jensen Shannon divergence*, *Chaos* **27**, 083118 (2017).
- [14] S J Sheather, *Density estimation* *Stat. Sci.* **19**, 588 (2004).
- [15] A Papoulis, *Probability, random variables and stochastic processes*, McGraw-Hill, New York (1991).
- [16] J Burbea, C R Rao, *On the convexity of some divergence measures based on entropy functions*, *IEEE T. Inform. Theory* **28**, 489 (1982).
- [17] J Lin, *Divergence measures based on the Shannon entropy*, *IEEE T. Inform. Theory* **37**, 145 (1991).
- [18] S Kullback, R A Leibler, *On information and sufficiency*, *Ann. Math. Stat.* **22**, 79 (1951).