

γ -Divergences

Leonardo E. Riveaud^{1,2}, Diego M. Mateos^{1,3,4,*}, and Pedro W. Lamberti^{1,5}

¹Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

²Facultad de Ingeniería, Universidad Nacional del Comahue (FAIN, UNComa)

³Facultad de Ciencia y Tecnología. Universidad Autónoma de Entre Ríos (UADER). Oro Verde, Entre Ríos, Argentina.

⁴Instituto de Matemática Aplicada del Litoral (IMAL-CONICET-UNL), CCT CONICET, Santa Fé, Argentina.

⁵Facultad de Matemática Astronomía y Física (FaMAF), Universidad Nacional de Córdoba. Córdoba, Argentina.

*Corresponding author: Diego M. Mateos, mateosdiego@gmail.com.

Abstract

Divergences have become a very useful tool for measuring similarity (or dissimilarity) between probability distributions. Depending on the field of application a more appropriate measure may be necessary. In this paper we introduce a family of divergences we call γ -divergences. They are based on the convexity property of the functions that generate them. We demonstrate that these divergences verify all the usually required properties, and we extend them to weighted probability distribution. We investigate their properties in the context of kernel theory. Finally, we apply our findings to the analysis of simulated and real time series.

Keywords— Divergences, Kernels, Time series analysis

1 Introduction

There exists a profusion of divergences between probability distributions. Remarkably when some of them are applied to the same statistical problem, in general they do not lead to indistinguishable results. Therefore, it is useful to have a large set of divergences. In general the divergences have different origins. Some are statistical, others originated in information theory and others borrowed from the realm of pure mathematics. The Fisher's metric is a conspicuous example of the first kind [1]; and the Kullback-Leibler is a well known example originated in information theory [2]. Measures of similarity (or dissimilarity), between probability distributions have become of great interest in physics (classical and quantum), biology, theory of multi layers networks and many other areas of science [3–6].

Not all divergence is adequate to treat every problem. So, having a variety of divergence could be useful for both pure theoretical studies and in the context of applications. Sometimes it has been possible to introduce families of divergences, labelling each member with a parameter [7, 8] or by giving a general structure depending on a function with certain characteristics. An example are the Csiszar's divergences or f -divergences.

Let $\Omega = \{\omega_1, \dots, \omega_N\}$ be a discrete sample space. The set of probability distributions on Ω can be identified with the simplex $\mathbb{P}^N = \{x^i \geq 0, i = 1, \dots, N, \sum_{i=1}^N x^i = 1\}$. Given two probability distributions $P = \{p_i\}_{i=1}^N$ and $Q = \{q_i\}_{i=1}^N$ belonging to $\mathbb{P}^N$, a Csiszar's divergence is defined by [9]

$$\mathcal{D}_f(P||Q) = \sum_i f\left(\frac{p_i}{q_i}\right) q_i$$

where $f(x)$ is a convex function such that $f(1) = 0$. This family has been extensively studied in the context of information geometry [10]. A remarkable result is that when p_i and $q_i = p_i + \delta p_i$ are two close probability distributions, the divergence $\mathcal{D}_f(P||Q)$ is proportional to the Fisher (Riemannian) metric:

$$\mathcal{D}_f(P, P + \delta P) \sim \frac{f''(1)}{2} \sum_i \frac{\delta p_i^2}{p_i} \quad (1)$$

A Csiszar' divergence is symmetric if and only if the function $f(u)$ satisfies:

$$uf\left(\frac{1}{u}\right) = f(u) + \beta(u - 1)$$

with β a constant.

The above mentioned Kullback-Leibler divergence corresponds to $f(t) = t \log t$ and it is not symmetric and the Jensen-Shannon divergence it is obtained by taking $f(t) = (t + 1) \log(\frac{2}{t+1}) + t \log t$, which clearly is symmetric [11].

In this work we present a new family of divergences that we call γ -divergences, based on the attributes of convex functions. By mimicking the structure of Euclidean metric, we propose a family of divergences that verify the basic properties of a "good" divergence [12].

This paper is organized as follows. In section 2 we introduce the family of distance based on convex functions. We demonstrate that these distances meet the requirements to be considered as divergences. Later on, we show the characteristics that this γ -divergence, and also extended our divergence to weighted and N -dimensional distribution. In section 3 we sketch a mode to check if a γ -divergence is a true metric. These results also provide a way to extend the γ -divergence to the realm of quantum mechanics. In section 4 we applied the divergence for the detection of dynamics changes in generated sequences and electroencephalographic signals. Finally, in section 5 we discussed the results obtained and future works are proposed.

2 The γ -Divergence family

The concept of metric space has a long and fruitful history. It dates back to the works of M. Fréchet and F. Hausdorff at the beginning of twenty century. Let us recall that a metric space is a pair (X, d) , where X is a topological space and $d : X \times X \rightarrow \mathbb{R}$ such that:

1. $d(x, y) \geq 0$ for every $x, y \in X$ and the equality is valid if and only if $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, y) + d(y, z) \geq d(x, z)$ for every $x, y, z \in X$

Obviously the idea of a metric space is inspired in the Euclidean space. In the context of measures of dissimilarity between probability distributions conditions 2 and 3 are not always satisfied. A long studied problem concerns the properties that a map $\phi : X \rightarrow X$ should have in order that the space $(X, \phi(d))$ be a metric one. For example it is well known that if (X, d) is a metric space, $(X, \frac{d}{d+1})$ and (X, d^α) with $\alpha \in (0, 1)$ also satisfy the metric properties.

We will focus our attention on $X = \mathbb{P}^N$. Let $P = \{p_i\}_{i=1}^N$ and $Q = \{q_i\}_{i=1}^N$ two discrete probability distributions belonging to $\mathbb{P}^N$. The square of the Euclidean distance between these two distributions can be written in the form:

$$\mathcal{E}(P||Q) = \sum_i (q_i - p_i)^2 = 2 \sum_i q_i^2 + 2 \sum_i p_i^2 - (p_i + q_i)^2 \quad (2)$$

or in an equivalent form:

$$\frac{1}{2} \mathcal{E}(P||Q) = \sum_i \left(q_i g(q_i) + p_i g(p_i) - (p_i + q_i) g\left(\frac{p_i + q_i}{2}\right) \right) \quad (3)$$

where $g(x) = x$ is the identity function. The space $(\mathbb{P}^N, \sqrt{\mathcal{E}})$ is a metric space.

The above mentioned Jensen-Shannon divergence (JSD), can be written in the form:

$$\mathcal{D}_{JS}(P||Q) = \sum_i \frac{1}{2} p_i \log(p_i) + \frac{1}{2} q_i \log(q_i) - \frac{1}{2} (p_i + q_i) \log\left(\frac{q_i + p_i}{2}\right) \quad (4)$$

which has the same structure that (3), with $g(x) = \frac{1}{2} \log x$. A remarkable property of the JSD is that its square root, $d_{JS} = \sqrt{\mathcal{D}_{JS}}$, it is a metric [12]. In both examples the products $x \cdot g(x)$ are convex functions. In some sense we can think the JSD as a “deformation” through the function $g(x)$ of the euclidean metric.

This simple observation leads us to propose, for each function $g(x)$ such that $x \cdot g(x)$ is convex, a divergence

$$\mathcal{D}_\gamma(P||Q) = 2 \sum_i \gamma_g(p_i, q_i) \quad (5)$$

with

$$\gamma_g(p_i, q_i) = \frac{1}{2} p_i g(p_i) + \frac{1}{2} q_i g(q_i) - \frac{1}{2} (p_i + q_i) g\left(\frac{p_i + q_i}{2}\right) \quad (6)$$

We will call from now on to $\mathcal{D}_\gamma(P||Q)$ the γ divergence.

The main properties of the γ divergence are presented in the following

Theorem: Let $P = \{p_i\}_{i=1}^N$ and $Q = \{q_i\}_{i=1}^N$ two probability distributions for a given N -state random variable X . Let $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $G(x) := x \cdot g(x)$ is a convex function. Then the functional defined in (5) satisfies the properties:

1. $\mathcal{D}_\gamma(P||Q) = \mathcal{D}_\gamma(Q||P)$ (*Symmetry*)
2. $\mathcal{D}_\gamma(P||Q) > 0$ for $Q \neq P$ (*Positivity*)
3. $\mathcal{D}_\gamma(P||Q) = 0 \iff P \equiv Q$

Proof: The symmetric character of $\mathcal{D}_\gamma(P||Q)$ is direct from the definition.

Let us recall that for a convex function $G(x)$ and for all $t \in [0, 1]$ and $p, q \in [0, 1]$ the Jensen inequality reads

$$t G(q) + (1 - t) G(p) \geq G(t q + (1 - t) p) \quad (7)$$

Therefore the convexity of $G(x) = x g(x)$ and the Jensen inequality for $t = 1/2$ imply that each term $\gamma_g(p_i, q_i)$ is positive and then $\mathcal{D}_\gamma$ is positive also.

Obviously $\mathcal{D}_\gamma(P||Q) = 0$ if $P \equiv Q$. On the contrary if $\mathcal{D}_\gamma(P||Q) = 0$ then $P = Q$. Indeed, for y fixed, the equation

$$0 = x g(x) + y g(y) - (x + y) g\left(\frac{x + y}{2}\right)$$

has only one solution at $x = y$.

◇

2.1 Similarity

Let us investigate the relation between the Csiszar's divergences and the γ -divergencies. Let $q_i = p_i + \delta p_i$ for each i , with $\sum_{i=1}^N \delta p_i = 0$, and let $g(x)$ two times differentiable. In this case it is direct evaluate:

$$\mathcal{D}_\gamma(P||P + \delta P) = \frac{1}{4} \sum_i [2g'(p_i) + g''(p_i)] \delta p_i^2 \quad (8)$$

where $g'(x) := \frac{dg}{dx}$.

If we require that at second order this approximation agree with the expansion of a Csiszar's divergences at same order, eq. (1), it must be valid that:

$$pg''(p) + 2g'(p) = \frac{f''(1)}{2p}$$

that is:

$$\frac{d}{dp}(p^2 g'(p)) = \frac{f''(1)}{2}$$

By integrating this equation, we get

$$pg(p) = G(p) = \frac{f''(1)}{2} p \log p + \alpha p \quad (9)$$

where α is a constant. The linear term does not contribute to the γ -Divergence. This make us to conclude that the only γ -Divergence that is a Csiszar's one is the JSD. This is a very important result, due that it enables to study to the γ -Divergence as a new class of divergences.

2.2 Weighted γ -Divergences

In several contexts, it is be useful to assign different relevance to each probability distribution, for example in Bayesian inference. This motivates us to propose a way to define a generalised γ -Divergence between weighted probability distribution. In expression (5) we can interpret that each probability distribution has assigned a weight $\frac{1}{2}$. Then, if we assign a weight π_P to distribution $P = p_i$ and a weight π_Q to the distribution $Q = q_i$, with $\pi_P + \pi_Q = 1$. we can introduce a weighted γ -Divergence in the form:

$$\mathcal{D}_\gamma^{\pi_P, \pi_Q}(P||Q) = 2 \sum_i \pi_P p_i g(p_i) + \pi_Q q_i g(q_i) - m_i g(m_i) = \quad (10)$$

$$= 2 \sum_i \pi_P p_i (g(p_i) - g(m_i)) + \pi_Q q_i (g(q_i) - g(m_i)) \quad (11)$$

where $m_i = \pi_P p_i + \pi_Q q_i$. This assignment assures that $\mathcal{D}_\gamma^{\pi_P, \pi_Q}(P||Q) \geq 0$, since

$$\gamma_g^{\pi_P, \pi_Q}(p_i, q_i) = \pi_P p_i g(p_i) + \pi_Q q_i g(q_i) - m_i g(m_i) \geq 0 \quad \forall i : 1, \dots, N \quad (12)$$

It is also possible to generalize the γ -Divergence for more than two probability distributions. Again we resort to the Jensen's inequality. Let $G(x)$ a real convex function and let $\{x_1, \dots, x_K\}$ be K real numbers belonging to the interval $[0, 1]$. Then

$$G\left(\sum_{k=1}^K \pi_k x_k\right) \leq \sum_{k=1}^K \pi_k G(x_k) \quad (13)$$

where $\pi_k \in [0, 1]$ such that $\sum_{k=1}^K \pi_k = 1$. Then by identifying $G(x) := x \cdot g(x)$, we have

$$\sum_{k=1}^K \pi_k x_k g(x_k) - \left(\sum_{k=1}^K \pi_k x_k\right) g\left(\sum_{k=1}^K \pi_k x_k\right) \geq 0 \quad (14)$$

We can interpret the left side of this inequality as an extension of the function γ_g previously defined, but now for K probability distributions:

$$\mathcal{D}_\gamma^{\pi_1, \dots, \pi_K}(P^1, \dots, P^K) := \sum_{i=1}^N \sum_{k=1}^K \pi_k P_i^k g(P_i^k) - \sum_{i=1}^N m_i g(m_i) \quad (15)$$

with $m_i := \sum_{k=1}^K \pi_k P_i^k$. We are assuming that the sample space is N dimensional and P_i^k represents the probability of occurrence of event i according to the distribution P^k .

3 Kernel properties of the γ -Divergences

Let Z be an arbitrary set. A function $\varphi : Z \times Z \rightarrow \mathbb{R}$ is called a negative definite kernel if is symmetric and for any $n \geq 2$, $\{z_1, \dots, z_n\} \subseteq Z$, and $\{c_1, \dots, c_n\}$ a collection of real numbers such that $\sum_{i=1}^n c_i = 0$, the inequality [13]

$$\sum_{j,k}^n c_j c_k \varphi(z_j, z_k) \leq 0 \quad (16)$$

is valid.

It is easy to show that the kernel associated to the euclidean distance

$$\varphi(x, y) = x^2 + y^2 - (x + y) \frac{x + y}{2}$$

is negative definite. It is possible also to prove that the kernel associate to the JSD is negative definite too.

The condition of being negative definite it is very relevant because a theorem by I. Schoenberg affirms that (Z, d) is a metric space if the square d^2 is a negative definite kernel.

The general structure of a γ -Divergence is a sum of terms of the form

$$\varphi(x, y) = xg(x) + yg(y) - (x + y)g\left(\frac{x + y}{2}\right) \quad (17)$$

If we were able to prove that $\varphi(x, y)$ is a negative definite kernel, then we could assure that $(\mathbb{P}^N, \mathcal{D}_\gamma^{\frac{1}{2}})$ is a metric space. From the general structure of the γ -Divergences, they will be true metrics if under the conditions of the Schoenberg theorem the inequality

$$-\sum_{j,k}^n c_j c_k (z_j + z_k) g\left(\frac{z_j + z_k}{2}\right) \leq 0$$

is valid.

Another consequence of the Schoenberg's theorem is that if inequality (eq.16) is satisfied for the kernel (eq.17), the metric space $(\mathbb{P}^N, \mathcal{D}_\gamma^{\frac{1}{2}})$ can be embedded into a Hilbert space $\mathcal{H}$

$$\Phi : \mathbb{P}^N \rightarrow \mathcal{H}$$

and

$$\|\Phi(P) - \Phi(Q)\| = \mathcal{D}_\gamma^{\frac{1}{2}}(P, Q)$$

with $\|\cdot\|$ is the norm induced by the Hilbert's space inner product. This is a very significant result if we are interested in extending the γ -Divergence to the realm of quantum mechanics. Worth mentioning that in the case of euclidean metric the embedding Φ induce the Wootter's distance between states $|\Psi_1\rangle$ and $|\Psi_2\rangle$ [14]:

$$d_W(|\Psi_1\rangle, |\Psi_2\rangle) = \arccos |\langle \Psi_1 | \Psi_2 \rangle|$$

4 Applications

In this section we use γ -divergences as a tool for detecting dynamical changes in artificially generated sequences and real EEG signals.

4.1 Dynamics changes detection

We use a Monte Carlo simulation to study the efficiency of γ -divergences to detect dynamical changes in binary sequences artificially generated [15]. For this purpose, we generated M binary sequences composed by two sub-sequences $s^i = s_1^i + s_2^i$ for $i = 1, \dots, M$ with length $L = L_{s_1} + L_{s_2}$. Each sub-sequence had a probability distribution $P_{s_1} = [p_{s_1} \ 1 - p_{s_1}]$ and $P_{s_2} = [p_{s_2} \ 1 - p_{s_2}]$ (Fig. 1A). We defined a pointer ρ which moves step by step across the sequences. In each step, we took two sequences with the same length (L_{win}), one on the right (s_r) and other on left (s_l) of ρ (Fig. 1B). We estimated the probability distribution for both sequences P_{s_l} and P_{s_r} , and calculated the γ -divergence between then $D_\gamma = D_\gamma(P_{s_l} || P_{s_r})$. This procedure was repeated for each $\rho \in [L_{win}+1 \ L - L_{win}]$ (Fig. 1C). The point (ρ^*) where the divergence reaches the maximum value $D_\gamma(\rho^*) = D_{\gamma_{max}}$ was the transition point between the sequence s_1 to s_2 . We did this analysis for the M generated sequences and then calculated their mean value and standard deviation. Beyond that this was developed for binary sequences, can be applied to all type of discrete or continuous (with previous quantification) sequences.

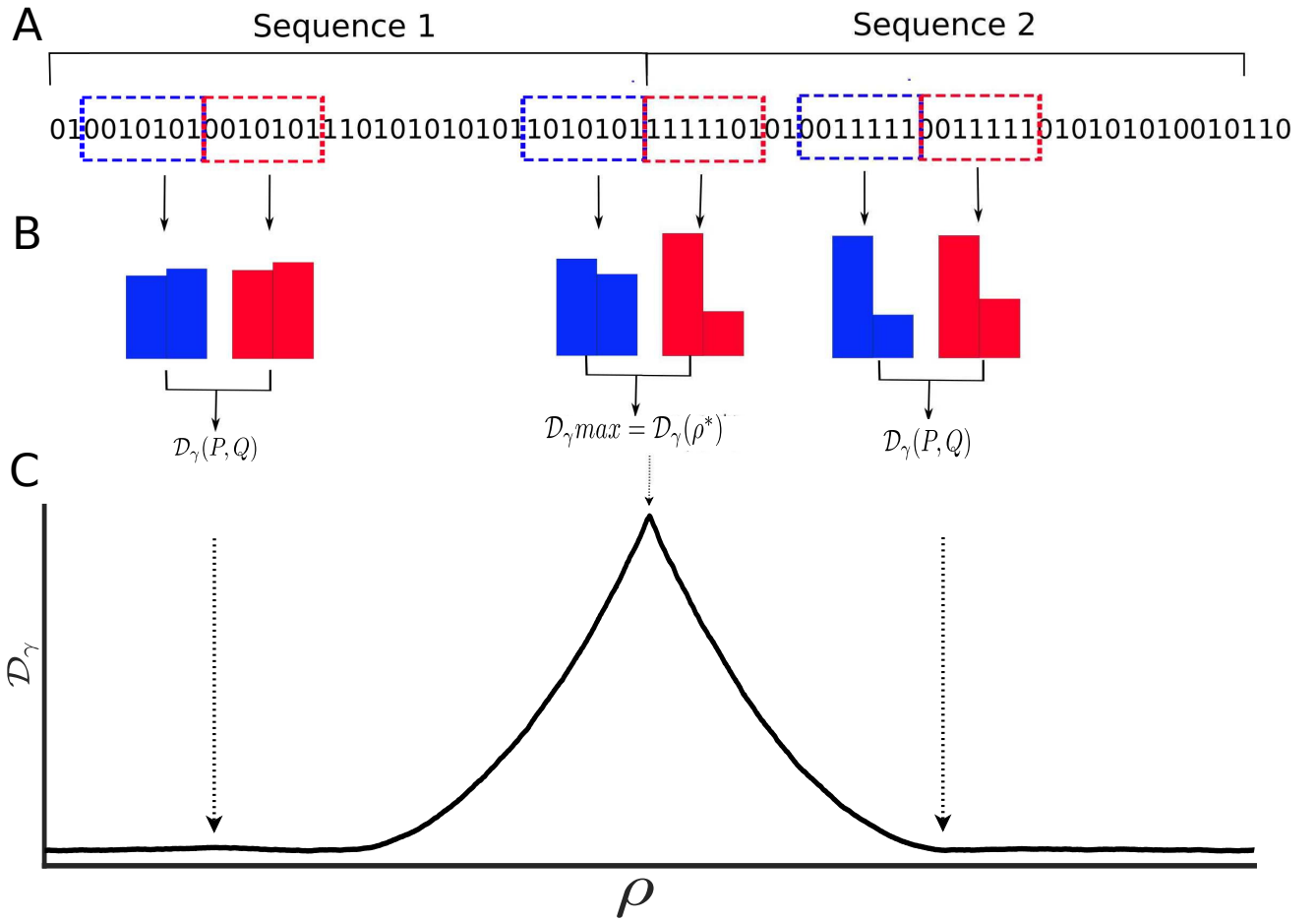


Figure 1: Window method used to detect distribution changes in a sequence. A) Generation of M combined sequences of $s = s_1 + s_2$, with s_1 probability distribution equal to P_{s_1} and s_2 to P_{s_2} . B) The pointer (ρ) moves over the sequence. Two subsequences are taken, on the left s_l (blue) and on the right s_r (red) of the ρ position. Then the probability distribution was calculated for each subsequences and γ -divergence is calculated. C) The same procedure is performed for all values of $\rho \in [L_{win}+1, L - L_{win}]$; the point (ρ^*) at which $D_\gamma = D_{\gamma_{max}}$ determines the transition from one sequence to another.

Figure 2 shows the γ -divergence for four specific $g(x)$ applied over a combined binary sequence. The first $s_1 = 3000$ points was generated with a probability $P = [0.5 \ 0.5]$ and the following $s_2 = 3000$ points with $Q = [0.4 \ 0.6]$. The analysis was made using the function ($g(x) = e^x, \log(x), \sqrt{x}, \sinh(x)$) all of

which satisfies the formal condition given in the theorem 2. We used a windows length of $L_{win} = 1000$ data point. The fill line represents the D_γ mean value, and the shadows are the standard deviation over the $M = 1000$ realization. Vertical dash line determines the point where sequence 1 and 2 are joined. For a better visualization of the results, the domain of the function goes from the domain $[1001 \ 5000]$ of the original sequence since the first and last 1000 points the D_γ are zero. We can see for all function $g(x)$ that the maximum divergence ($D_{\gamma_{max}}$) is reached at the exact point where the sequence changes the probability distribution. In all cases, maximum divergence values are much higher than the standard deviations, demonstrating its statistical significance.

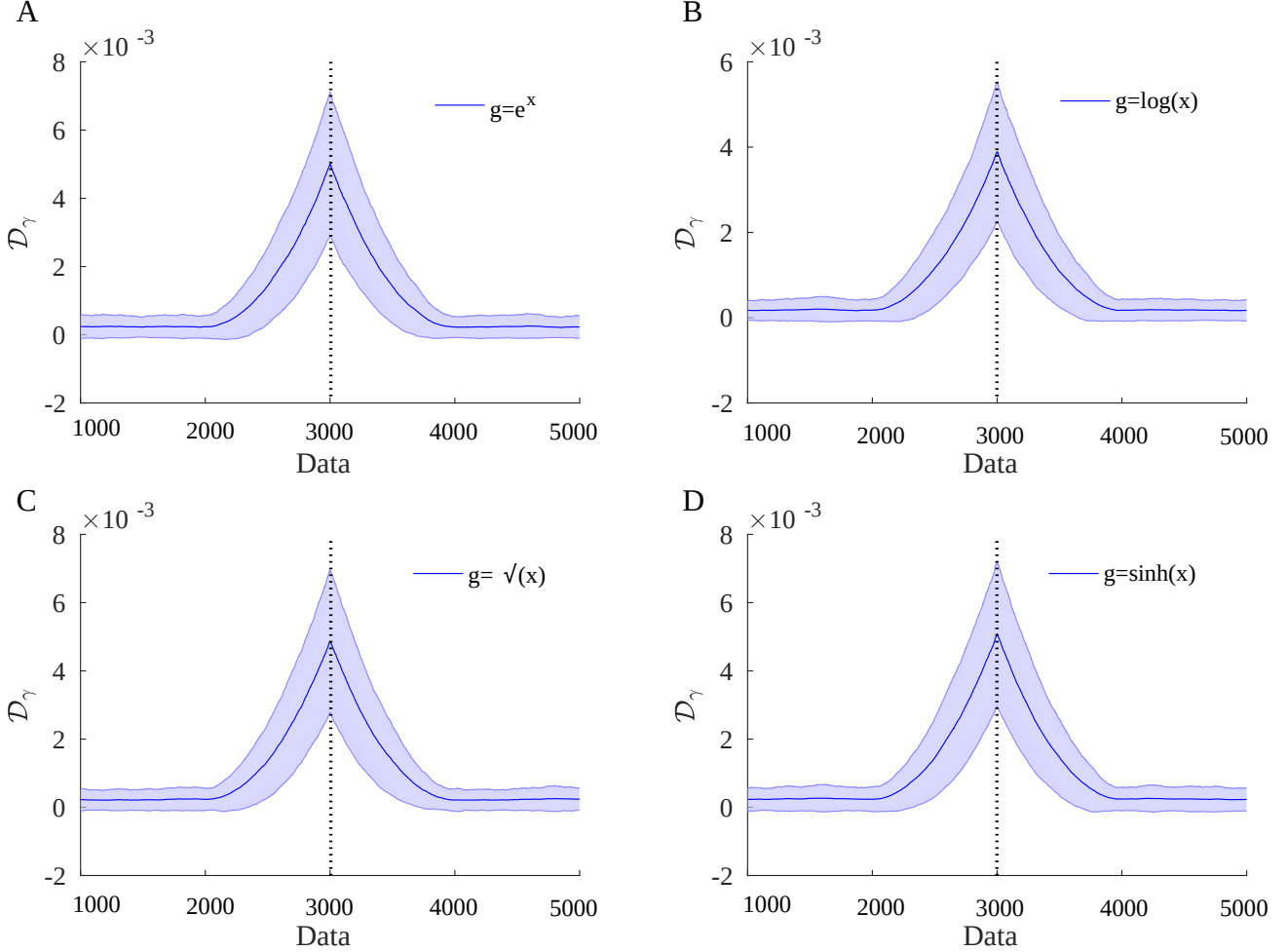


Figure 2: γ -divergence analysis over a combined binary sequence. The first 3000 points correspond to s_1 with probability distribution $P = [0.5 \ 0.5]$ and the latter s_2 with probability distribution $Q = [0.4 \ 0.6]$. The windows length are $L_{win} = 1000$. The full line corresponds to the D_γ means that the values and the shadow represent the standard derivation over $M = 1000$ realization. The vertical line represents the exact point where the two sequences join. The same analysis was made for different function $g(x)$: A) e^x , B) $\log(x)$, C) $\sqrt{x}$ and D) $\sinh(x)$. In all cases the $D_{\gamma_{max}}$ is reached at the exact point where the two sequences join.

Then, we wanted to study the detection limit of the γ -divergence. In other words, for binary sequences we would like to saw what the smallest probability distribution difference was that can be detected by this family of divergences. To fulfil these aims we generated four combine binary sequences with probability distribution increasingly closer. We analyzed the sequences with the function ($g(x) = e^x, \log(x), \sqrt{x}, \sinh(x)$). The windows length used were $L_{win} = 1000$ data point. Fig. 3 shows the analysis for function $g(x) = e^x$.

As the two probability distribution (P and Q) become closer, the maximum divergence value ($D_{\gamma_{max}}$)

decreases, being in the limit of detection for $P = [0.5 \ 0.5]$ and $Q = [0.45 \ 0.55]$ (Fig. 3C) and, making it impossible to distinguish between them when $P = [0.5 \ 0.5]$ and $Q = [0.51 \ 0.49]$ (Fig. 3D). Similar results we obtained for the function $\log(x)$, $\sqrt{x}$, $\sinh(x)$.

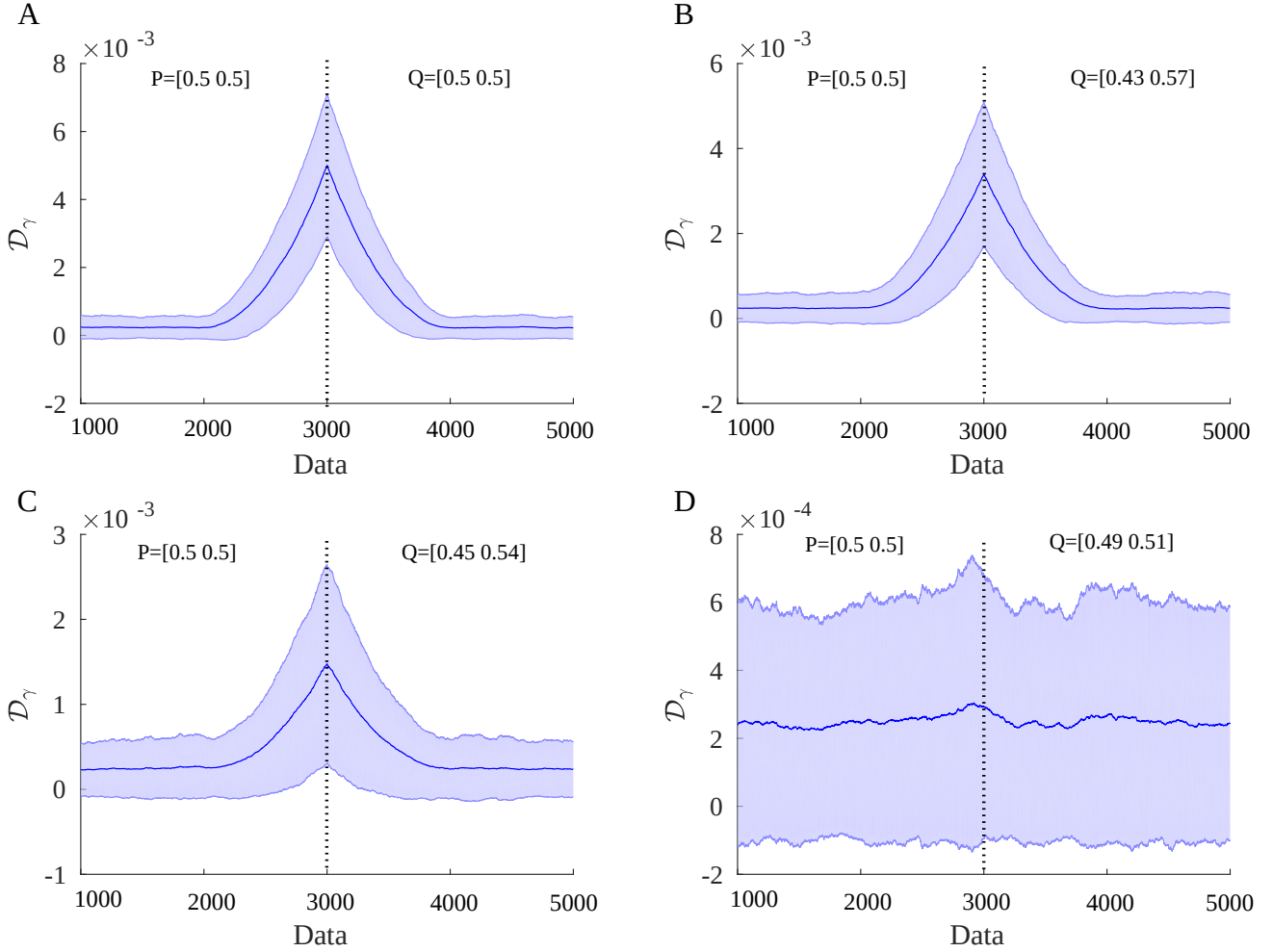


Figure 3: γ -divergence analysis over a combined binary sequence with different P and Q . The first $s_1 = 3000$ points have a P probability distribution and the latter $s_2 = 3000$ have Q probability distribution. The function used for the analysis was $g(x) = e^x$ and the windows length were $L_{win} = 1000$. The vertical line represents the exact point where the two sequences join. A,B As P and Q become more similar, the divergence maximum ($D_{\gamma_{max}}$) becomes smaller; however, detection remains significant. C) For $P = [0.5 \ 0.5]$ and $Q = [0.45 \ 0.54]$ detection between the two signals is observed, but the statistical significance lies in the limit. D) For $P = [0.5 \ 0.5]$ and $Q = [0.49 \ 0.51]$ The divergence can't distinguish between the two sequences.

4.2 Transition detection over EEG sleep recording

In the second example, we used the γ -divergence to identify the transition between sleep states in an electroencephalogram (EEG) signal from a sleeping patient. Sleeping is a dynamic activity, during which many processes are vital to health and well-being take place. It is essential to help to maintain mood, memory, and cognitive performance [16–18]. A specialist defines two primary sleep stages: *rapid eye movement* (REM) sleep and *non-REM* sleep (NREM). REM is defined as an active sleeping period showing an intense brain activity. Brain waves remain fast and desynchronized similar to those in the waking state. This is also the sleep stage is the state in which the majority of dreams are produced. In NREM sleep stage the physiological activity decrease, the brain waves get slower and have greater

amplitude, breathing, and heart rate slows down and blood pressure drops. The NREM phase is composed by three stages: N1, N2, and N3. The N1 stage is characterized by perceived drowsiness or transition from being awake to falling asleep observed by slowing down the brain waves and muscle activity. Stage N2 is a period of light sleep during which eye movement stops. Brain waves become slower (Theta waves (4-7 Hz)) with occasional bursts of rapid waves (12-14 Hz) called *sleep spindles*, coupled with spontaneous periods of muscle tone mixed. Lastly, stage N3 is characterized by the presence of Delta (0.5-4 Hz) slow waves, interspersed with smaller, faster waves [19]. In the N3 stage is a deep sleep stage, without eye movement, and a decrease of muscle activity, resembling a coma state. Usually, sleepers pass through these four stages (REM, N1, N2, and N3) cyclically. A complete sleep cycle takes an average of 90 to 110 minutes, with each one lasting between 5 to 15 minutes. These cycles must be maintained for healthy body function in awake state [20]. Developing tools that can detect the changes in dynamic sleep stages over EEG signal are highly essential for studying patients with sleep disorders [21]. We used our γ -divergence to detect changes in the dynamics of the EEG to allow us to distinguish between one state from another.

The data were taken from the *Physionet database: The Sleep-EDF Database [Expanded]* [22, 23], and are freely available at [24]. The EEG were recorded (Fpz-Cz) bipolar channel and the sampling frequency was 100 Hz. Initially, we extracted segments from the original signal belonging to the five different sleep states: Awake, REM, N1, N2, N3¹. Each segments had 6000 points (corresponding to 60 sec of recording) and were joined into a single signal² Fig. 4A. The signal was preprocessed with a band-pass filter between 0.5 – 60 Hz. We quantify the signal using the permutation vector method approach [25], with parameter $d = 4$ and $\tau = 1$. We applied the γ -divergence following the method used in the previous section for the binary sequence. The functions used were $(g(x) = e^x, \log(x), \sqrt{x}, \sinh(x))$.

Figure 4B shows that the divergence detects the transition between different sleep states for all the functions $g(x)$. The $\log(x)$ function shows the highest values and the best differentiation between stages.

The transition between Awake-REM and N2-N3 are more remarkable than in REM-N1 and N1-N2. There is a significant differentiation between N2-N3 because of N3 is the deepest sleep stage. In this stage the body becomes more insensitive to outside stimuli than seeming to coma state. In this state, the EEG signal is mostly composed by slow waves (Delta and Theta) causing the brain dynamics to be sharply different from the other states. Lower γ -divergences values were found between N1-N2 states showing that both states share similar characteristics in their dynamics. Particularly, N1 is characterized by drowsiness slowing down the brain waves and muscle activity. While N1 is a period of light sleep during which eye movement stops, the two stages (N1-N2) are very similar with the difference that in N1 occasional bursts of rapid waves (12-14 Hz) called *sleep spindles* appear.

A similar result can be found between REM and N1. REM and N1 also present lower values of divergence. This is expected considering that the EEG of REM sleep contains frequencies present in the “awake” state and in the lighter stages of sleep N1 [26]. Despite the similarities of the REM and N1, there are still enough differences between them such that statistically different values are obtained. The presence of 11-16 Hz activity (sleep spindles) in N1, and more abundant alpha activity (8-13Hz) in REM sleep means that these two stages present activity at an overlapping frequency range, which explains the proximity of the divergence values obtained. Difficulty in detecting N1 and REM sleep has also been found using other measures [27, 28].

5 Discussion

In the first place, we showed the existence of a close relationship between the square of the Euclidean metric and the Jensen-Shannon divergence establishing that both belong to the same family of functional. In some sense we can think to the JSD as a deformation of the euclidean distance. Based on this,

¹This was done based on the notes provided by the database

²This was done for a better visualization of the results, because the time of each sleep state can vary from seconds to several minutes.

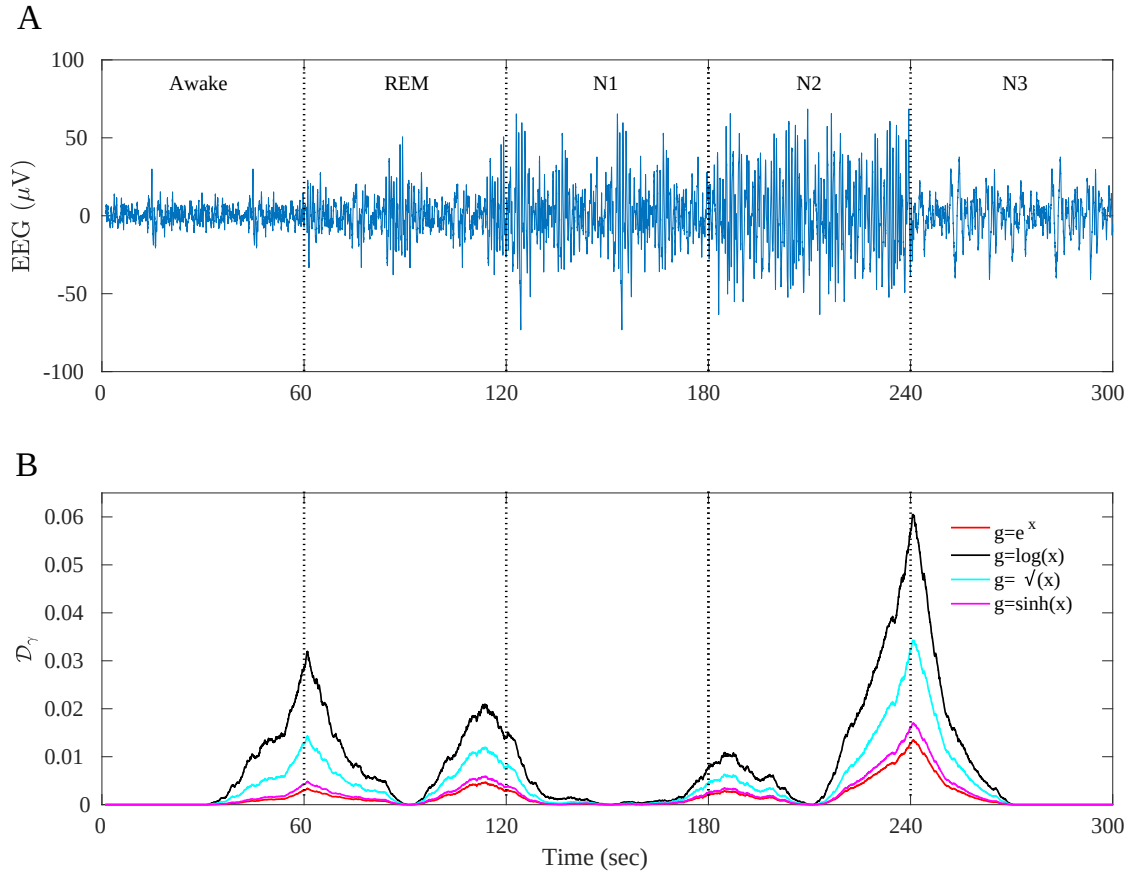


Figure 4: Application the γ -Divergence over a sleep EEG signal using a running windows. A) The EEG signal is composed by 5 sub-signal belonged different sleep stages (Awake, REM, N1, n2, n3). Each state has 6000 point corresponding to 60 sec recording (dashes horizontal lines). B) Using a running windows method the $\mathcal{D}_\gamma$ was applied taken four $g(x)$ compared for the study. The signal was quantified with the permutation vectors with parameter $d = 4$ and $\tau = 1$. For all functions, the maximum values $\mathcal{D}_{\gamma_{max}}$ were reached in the exact point where a transition between sleep states exists.

we introduced a family of divergences we called γ -divergence that depend on the properties of convex functions. Subsequently, we introduced generalized versions of these divergences (weighted versions).

We give an approach to the new divergences from the negative definite kernels theory, This allows us to explore an extension of the γ -divergence between quantum states. This connection will be explored extensively in a future work.

Finally, we applied the γ -divergence in simulated and real sequences. We showed that all the functions $g(x)$ could detect with high significance the exact point where sequences with different probabilities joined. Moreover, we studied the detection threshold –the point where the divergence can no longer distinguish between the two signals–. Later, we analysed EEG signal from sleep patient. We could detect the points where the signal changes its dynamics due to the change in the state of sleep. Showing it can be an alternative tool for detecting different sleep states.

It is also important to mention that when using divergences, we had consider the significance of the value obtained. This significance is what allows us to say if the distance found is really a real and not just a measure of the statistical fluctuations. In this work we used the standard deviation calculated over M realizations as a measure of significance. However, many times we do not have more than one realization being this method useless. Therefore, a theoretical study of the field is necessary. Some studies on this topic have already been addressed by Grosse et al. for the Jensen-Shannon divergence [15]. Similar studies should be carried out for this family of *gamma*-divergence. However, this exceeds this work but will be addressed in the near future.

6 Acknowledgements

We are grateful to Secretaria de Ciencia y Tecnica de la Universidad Nacional de Córdoba, Argentina, for financial assistance. This work was partially supported by Grant PIP 519/2019 from CONICET, Argentina.

References

- [1] R. A. Fisher. Theory of statistical estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 22, pages 700–725. Cambridge University Press, 1925.
- [2] T.M. Cover and J.A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
- [3] M. Parry and E. Fischbach. Probability distribution of distance in a uniform ellipsoid: Theory and applications to physics. *Journal of Mathematical Physics*, 41(4):2417–2433, 2000.
- [4] S.N. Ondimu and H. Murase. Effect of probability-distance based markovian texture extraction on discrimination in biological imaging. *Computers and Electronics in Agriculture*, 63(1):2–12, 2008.
- [5] A. Kostin. Probability distribution of distance between pairs of nearest stations in wireless network. *Electronics Letters*, 46(18):1299–1300, 2010.
- [6] A.P. Majtey, P.W. Lamberti, and D.P. Prato. Jensen-shannon divergence as a measure of distinguishability between mixed quantum states. *Physical Review A*, 72(5):052310, 2005.
- [7] T. M Osán, D.G. Bussandri, and P.W. Lamberti. Monoparametric family of metrics derived from classical jensen–shannon divergence. *Physica A: Statistical Mechanics and its Applications*, 495:336–344, 2018.
- [8] A.P. Majtey, P.W. Lamberti, and A. Plastino. A monoparametric family of metrics for statistical mechanics. *Physica A: Statistical Mechanics and its Applications*, 344(3-4):547–553, 2004.

- [9] I. Csiszár. Information measures: A critical survey. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pages 73–86, 1974.
- [10] J. Burbea and C. R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12(4):575–596, 1982.
- [11] S.M. Ali and S.D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.
- [12] D.M. Endres and J.E. Schindelin. A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860, 2003.
- [13] P. Ressel C. Berg, J. Christensen. *Harmonic Analysis on Semigroups*. Springer, 1984.
- [14] W.K. Wootters. Statistical distance and hilbert space. *Physical Review D*, 23(2):357, 1981.
- [15] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, and H.E. Stanley. Analysis of symbolic sequences using the jensen-shannon divergence. *Physical Review E*, 65(4):041905, 2002.
- [16] R.M. Benca, M. Okawa, M. Uchiyama, S. Ozaki, T. Nakajima, K. Shibui, and W.H Obermeyer. Sleep and mood disorders. *Sleep medicine reviews*, 1(1):45–56, 1997.
- [17] R. Stickgold. Sleep-dependent memory consolidation. *Nature*, 437(7063):1272–1278, 2005.
- [18] J.J. Pilcher and A.S. Walters. How sleep deprivation affects psychological variables related to college students’ cognitive performance. *Journal of American College Health*, 46(3):121–126, 1997.
- [19] R. Boostani, F. Karimzadeh, and M. Nami. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer methods and programs in biomedicine*, 140:77–91, 2017.
- [20] R. D. Ogilvie. The process of falling asleep. *Sleep Medicine Reviews*, 5(3):247–270, 2001.
- [21] G. Tononi and C. Cirelli. Sleep function and synaptic homeostasis. *Sleep medicine reviews*, 10(1):49–62, 2006.
- [22] A.L. Goldberger, L. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, RG Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. Physiobank, physiotoolkit, and physionet: Circulation. *Discovery*, 101(23):1, 1997.
- [23] B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C Kamphuisen, and J.J.L. Obery. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the eeg. *IEEE Transactions on Biomedical Engineering*, 47(9):1185–1194, 2000.
- [24] Physionet databanck. Sleep-EDF Database Expanded. \url{https://physionet.org/content/sleep-edfx/1.0.0/}.
- [25] C. Bandt and B. Pompe. Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 88(17):174102, 2002.
- [26] N. Nicolaou and J. Georgiou. The use of permutation entropy to characterize sleep electroencephalograms. *Clinical EEG and Neuroscience*, 42(1):24–28, 2011.
- [27] Q. Noirhomme, M. Boly, V. Bonhomme, P. Boveroux, C. Phillips, P. Peigneux, Soddu, and Others. Bispectral index correlate with regional cerebral blood flow during sleep. *Archives Italiennes de Biologie*, 147(1/2):51–57, 2009.
- [28] D.M. Mateos, J. Gómez-Ramírez, and O.A. Rosso. Using time causal quantifiers to characterize sleep stages. *bioRxiv*, page 550152, 2019.