Analysis of families of divergences to compare Gaussian processes modeled by sums of complex exponentials disturbed by additive white noises

Eric Grivel

Version of Record: https://www.sciencedirect.com/science/article/pii/S1051200422000537
Manuscript_025451f128523594438de3eef6e1866c

Analysis of families of divergences to compare Gaussian processes modeled by sums of complex exponentials disturbed by additive white noises

Eric Grivel[a]

[a]*Bordeaux University - INP Bordeaux - IMS - UMR CNRS 5218, FRANCE*

Abstract

The purpose of this paper is first to derive the expressions of various divergences that can be expressed from the Chernoff coefficient in order to compare two probability density functions of vectors storing k consecutive samples of a sum of complex exponentials disturbed by an additive white noise. This includes the Chernoff divergence and the α-divergence for instance. Tsallis, reversed Tsallis and Sharma-Mittal divergences are also addressed as well as the β-, γ- and $\alpha\gamma$-divergences. The behaviors of the divergences are studied when k increases and tends to infinity. Depending on the divergence used, the divergence rate or the asymptotic normalized increment is considered. Expressions that encompass the divergence rate or the asymptotic normalized increment of the divergences are also given. Comments and illustrations to compare random processes are then given. This study makes it possible to show the advantages of the Kullback-Leibler divergence when studying this type of process.

Keywords: Kullback-Leibler divergence, Chernoff coefficient, divergence rate, process comparison, sum of noisy complex exponentials.

1. Introduction

Random processes can be modeled by different types of models depending on their statistical properties in terms of correlation and spectrum. For instance, in some speech enhancement applications, the signal can be modeled either by an autoregressive process or by a sum of complex exponentials, both disturbed by an additive white noise. Depending on the model chosen, Kalman and particle filters or subspace methods can be used to estimate the signal from the noisy observations. In spectrum analysis, some approaches based on an *a priori* model can be considered. Once again, the models mentioned above can be considered. In this paper, we will focus our attention on one of these models. It is defined by a set of parameters, namely the normalized angular frequencies, the variances of the magnitudes of the complex exponentials and the variance of the additive noise.

When dealing with statistical change detection[1] or time series comparison, different approaches exist: comparing the model parameters, using a spectral distance or exploiting a divergence. Divergences are not only used in this case. Indeed, a great deal of interest has been paid to the divergences by various researchers working in different fields: signal and image processing but also statistics and information theory. See for instance [4, 12, 25, 27, 28, 41]. Besides the use of divergences in various applications, different topics dealing with the divergences have been addressed for the last years. Among the recent theoretical contributions, some are highlighted below:

1/ Authors like Nielsen [31–34] presented closed-form expressions of divergences for some families of distributions or studied new generalizations of existing divergences [10, 22, 23, 32, 33, 44].

2/ The estimation of the divergence, when the data are Gaussian or not, is another issue addressed by some researchers [8, 9, 30, 34, 42]. Indeed, when

[1] The purpose is to detect when the statistical properties of the signal evolve. To this end, the beginning of the signal serves as a reference and it is compared with another frame of the signal, using a sliding window.

dealing with Gaussian processes, the mean and the covariance matrices could be estimated from the data, but the number of samples available has a strong impact on the accuracy of the estimates. The estimations of the Kullback-Leibler (KL) divergence between two probability density functions (pdfs) that are not necessarily Gaussian by using sets of data were studied in [8, 9] for instance. When considering the KL importance estimation procedure (KLIEP), the density ratio is estimated instead of estimating the distributions separately. The reader may refer to the works proposed by Sugiyama *et al.* [42, 43, 45]. It should be noted that Nguyen *et al.* [30] also recall different works dealing with the estimation of the divergence in the introduction of their paper before presenting M-estimators for the KL divergence and the density ratio.

3/ In information theory, there are various quantities such as the entropy, the conditional entropy, the mutual information, the cross-entropy and the divergence. Very often, having information of this quantity per unit of time can be of interest. It characterizes a growth rate. Thus, the entropy rate for a stochastic process is the limit (when it exists) of the joint entropy of n samples divided by n, when n tends to infinity. Depending on the properties of the process, it can be related to the asymptotic equipartition property. When studying an independent identically distributed (i.i.d.) sequence, the corresponding Shannon and Rényi entropy rates are the entropy of the distribution. When dealing with a Markov chain respecting some conditions, the entropy rate depends on the stationary distribution and the transition probabilities. In [35], Rached *et al.* showed that the rate of the Rényi entropy of order α is related to the largest positive eigenvalue of the matrix whose elements are the transition probabilities at the power α. Formulas of entropy rates for (h, ϕ)-entropy and ways to estimate them are presented in [7]. It should be noted that the entropy rates are still the core of research activities -See for instance [38] [15]- and can be used in very different cases such as the characterization of complex networks in the real world [16] or the analysis of the voice pathology evolution [40]. For the last years, a great deal of interest has been paid to these "rates" and not only to the entropy rates. Thus, in [36] and [37], the authors respectively de-

rived a formula of the Rényi and KL divergence rates when comparing two time-invariant finite-alphabet Markov sources of arbitrary order. Moreover, the divergence rates for stationary Gaussian processes was studied initially in [20] and then in [13]. In these papers, the authors respectively provide the expression of the Kullback-Leibler (KL) divergence rate and the Rényi divergence rate for zero-mean Gaussian processes by using the theory of Toeplitz matrices and the properties of the asymptotic distribution of the eigenvalues of Toeplitz forms. In [17], we recently show that using the KL divergence rate is equal to the Itakura–Saito distance, up to a multiplicative factor when dealing with Gaussian stationary ARMA processes.

The Kullback-Leibler (KL) divergence [24], related to the Shannon entropy, is very popular but there are many other divergences that have been proposed in the literature and many of them correspond to a generalization of the KL divergence. See [2] for instance and figure 1.

In this paper, we propose to analyze how the divergences based on the Chernoff coefficients between the pdfs of vectors storing k consecutive samples of Gaussian processes modeled by sums of complex exponentials disturbed by white noises (NSCE for noisy sum of complex exponentials) evolve when k increases. Among them, the Chernoff, Bhattacharyya and Rényi divergences [3, 5, 39] are first investigated. Then the α-divergences [1] that include the squared Hellinger distance are studied. We also analyze the Tsallis, reversed Tsallis and Sharma-Mittal divergences. Finally, the β-divergence, the γ-divergence and the $\alpha\gamma$-divergences are considered [11] [6]. In each case, we provide the analytical expression of the divergences when dealing with Gaussian processes. Then, we analyze how the divergences evolve when the number of variates k increases. In some cases, this makes it possible to deduce what is called the "divergence rate", which corresponds to the asymptotic increment of the divergence, *i.e.* the difference between the divergences computed for $k+1$ and k variates, when k consecutive samples and tends to infinity. This means that the divergence as a function of k tends to an asymptote, the slope of which is the divergence rate. In other cases, we rather focus our attention on the asymptotic normalized

increment, *i.e.* the difference between the divergences computed for $k+1$ and k variates which is then divided by the divergence computed for k variates, when k tends to infinity. We will see whether these quantities depend on all the parameters characterizing the sums of complex exponentials disturbed by an additive noise, or not. This will help us to conclude whether they can be useful to compare random processes or to detect statistical change on data over time. This work is hence a follow up of previous papers [19] [17] that dealt with the behaviours of the Jeffreys and Rényi divergences.

The remainder of the paper is organized as follows: after recalling the properties of Gaussian processes modeled by sums of complex exponentials disturbed by white noises, the definitions of the divergences are recalled as well as the links they have between them. Then, the Gaussian case is addressed. The expressions of the divergences are given for this particular choice of pdf. We analyze how the increments of the divergences between pdfs of random vectors storing k consecutive samples of NSCE processes evolve when k increases and tends to infinity. They correspond to the divergence rate. When dealing with the Chernoff, Bhattacharyya, Rényi, γ- and $\alpha\gamma$-divergences, we provide a general expression that encompasses all the expressions of the divergence rates. When studying the α-divergences, the Tsallis, the reversed Tsallis and Sharma-Mittal (SM) divergences, a general expression of the normalized increment is provided and studied. Some illustrations as well as comments are then given. Three appendices are provided. The first one makes it possible to obtain the analytical expressions of integrals of products of Gaussian pdfs at different powers. This can be useful to get the analytical expression of the β-divergence and the γ-divergence in the Gaussian case. The second one provides another way to get the result in the Gaussian case based on [31]. The last one gives the mathematical details showing the link between the Rényi divergence and the KL divergence when dealing with NSCE processes.

2. Some properties about the processes to be compared

In this paper, we propose to focus our attention on a class of random processes that can be modeled by a sum of complex exponentials disturbed by an additive white noise. This type of modeling has been often used in the field of spectrum analysis when dealing with MUSIC and ESPRIT[2]. This model is an illustration of the Wold decomposition because the sum of exponentials characterizes a discrete spectrum while the additive white noise has a continuous spectrum [29].

Let us assume that the i^{th} random process, with $i = 1, 2$, is modeled by a sum of M_i zero-mean exponentials as follows:

$$\begin{aligned} x_i(n) &= \sum_{m=1}^{M_i} A_{m,i} e^{jn\theta_{m,i}} \\ &= \begin{bmatrix} e^{jn\theta_{1,i}} & \dots & e^{jn\theta_{M_i,i}} \end{bmatrix} \begin{bmatrix} A_{1,i} & \dots & A_{M_i,i} \end{bmatrix}^T \end{aligned} \tag{1}$$

where n refers to the n^{th} sample, $j^2 = -1$, $\{\theta_{m,i}\}_{m=1,\dots,M_i}$ are the normalized angular frequencies in the interval $[-\pi, \pi)$. For each process, they differ from each other. However, the two processes to be compared can have common normalized angular frequencies. In addition, $\{A_{m,i}\}_{m=1,\dots,M_i}$ are the zero-mean Gaussian random magnitudes with variances $\{\gamma_{m,i}\}_{m=1,\dots,M_i}$. These magnitudes are uncorrelated with each other.

The Toeplitz covariance matrix $Q_{SCE,k,i}$ of size $k \times k$ of the column vector storing k consecutive samples of the process x_i is then given by:

$$Q_{SCE,k,i} = S_{k,i} \Phi_i S_{k,i}^H \tag{2}$$

[2]ESPRIT stands for "estimation of signal parameter by rotational variance technique" whereas MUSIC means "multiple signal classification". They are widely used in spectrum analysis and array processing.

In (2), $S_{k,i}$ is a matrix of size $k \times M_i$ that stores M_i column vectors as follows:

$$S_{k,i} = \begin{bmatrix} S_{k,i}^1 & \dots & S_{k,i}^{M_i} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\theta_{1,i}} & \dots & e^{j\theta_{M_i,i}} \\ \vdots & & \vdots \\ e^{j(k-1)\theta_{1,i}} & \dots & e^{j(k-1)\theta_{M_i,i}} \end{bmatrix} \tag{3}$$

In addition, Φ_i is a diagonal matrix of size $M_i \times M_i$ whose main diagonal is defined by the variances $\{\gamma_{m,i}\}_{m=1,\dots,M_i}$ of the magnitudes $\{A_{m,i}\}_{m=1,\dots,M_i}$.

Remark. *The vectors $\{S_{k,i}^m\}_{m=1,\dots,M_i}$ have the following properties that will be useful in the rest of the paper:*

$$\begin{cases} \frac{1}{k}\left(S_{k,i}^m\right)^H S_{k,i}^m & = 1 \\ \frac{1}{k}\left(S_{k,i}^m\right)^H S_{k,i}^n & = \frac{sin(\frac{k(\theta_{m,i}-\theta_{n,i})}{2})}{ksin(\frac{(\theta_{m,i}-\theta_{n,i})}{2})} e^{j\left(\frac{(k-1)(\theta_{m,i}-\theta_{n,i})}{2}\right)} \end{cases} \tag{4}$$

Since $\lim_{k\to+\infty} |\frac{sin(\frac{k(\theta_{m,i}-\theta_{n,i})}{2})}{ksin(\frac{(\theta_{m,i}-\theta_{n,i})}{2})}| = 0$ for any pair $\theta_{n,i} \neq \theta_{m,i}$, the asymptotic properties of the vectors $\{S_{k,i}^m\}_{m=1,\dots,M_i}$ are:

$$\begin{cases} \frac{1}{k}\left(S_{k,i}^m\right)^H S_{k,i}^m = 1 \\ \lim_{k\to+\infty} \frac{1}{k}\left(S_{k,i}^m\right)^H S_{k,i}^n = 0 \end{cases} \tag{5}$$

The process defined above is then disturbed by an additive zero-mean white noise with variance σ_i^2, leading to a covariance matrix denoted as:

$$Q_{k,i} = Q_{SCE,k,i} + \sigma_i^2 I_k \tag{6}$$

where I_k is the identity matrix of size k.

Among the other properties that will be useful in the mathematical development that will follow, one can mention that the determinant of the covariance matrix of size $k+1$ can be expressed as[3]:

$$|Q_{k+1,i}| = \begin{vmatrix} Q_{k,i} & R_{k,i} \\ R_{k,i}^H & r_{0,i} \end{vmatrix} = r_{0,i} \times |Q_{k,i} - R_{k,i}\frac{1}{r_{0,i}}R_{k,i}^H| \tag{7}$$

[3]Let a matrix be defined as $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ where D is assumed to be invertible. It is post-

where $r_{0,i}$ is the covariance function of the i^{th} process for lag 0 while $R_{k,i}$ is a column vector storing the covariance function of the i^{th} process from lags k down to 1. Using the matrix determinant lemma[4], this leads to:

$$|Q_{k+1,i}| = |Q_{k,i}| \times r_{0,i} \times (1 - \frac{1}{r_{0,i}} R_{k,i}^H Q_{k,i}^{-1} R_{k,i}) \tag{8}$$

$$= |Q_{k,i}| \times (r_{0,i} - R_{k,i}^H Q_{k,i}^{-1} R_{k,i}) \tag{9}$$

Therefore, the ratio of the determinants of the covariance matrices of consecutive sizes satisfies:

$$\frac{|Q_{k+1,i}|}{|Q_{k,i}|} = r_{0,i} - R_{k,i}^H Q_{k,i}^{-1} R_{k,i} \tag{10}$$

Let us now look at the expression above when k increases and tends to infinity. To this end, on the one hand, using the matrix inversion lemma[5] on (6), one gets:

$$Q_{k,i}^{-1} = \frac{1}{\sigma_i^2} \left(I_k - S_{k,i} \left(\sigma_i^2 \Phi_i^{-1} + S_{k,i}^H S_{k,i} \right)^{-1} S_{k,i}^H \right) \tag{11}$$

In a previous paper [19], by applying again the matrix inversion lemma on the matrix $\left(\sigma_i^2 \Phi_i^{-1} + S_{k,i}^H S_{k,i} \right)$ and by using the asymptotic properties (5), it was

---

multiplied by $\begin{bmatrix} I & \mathbb{0} \\ -D^{-1}C & I \end{bmatrix}$, with I the identity matrix and $\mathbb{0}$ is a matrix of zeros, leading to the matrix $\begin{bmatrix} A - BD^{-1}C & B \\ \mathbb{0} & D \end{bmatrix}$, where $A - BD^{-1}C$ is the Schur complement. Taking the determinant and given the triangular structure of two of the three matrices, one has:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} \begin{vmatrix} I & \mathbb{0} \\ -D^{-1}C & I \end{vmatrix} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} A - BD^{-1}C & B \\ \mathbb{0} & D \end{vmatrix} = |A - BD^{-1}C||D|$$

[4] Assuming A invertible and v and v two column vectors, one has:

$$\begin{vmatrix} I & 0 \\ v^T & 1 \end{vmatrix} \begin{vmatrix} I + A^{-1}uv^T & A^{-1}u \\ 0 & 1 \end{vmatrix} \begin{vmatrix} I & 0 \\ -v^T & 1 \end{vmatrix} = \begin{vmatrix} I & A^{-1}u \\ 0 & 1 + v^T A^{-1} u \end{vmatrix}$$

Therefore, $|I + A^{-1}uv^T| = 1 + v^T A^{-1} u$ and $|A + uv^T| = |A|(1 + v^T A^{-1} u)$.

[5] Given A, U, C and V be matrices with A and C assumed to be invertible. One has:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U \left(C^{-1} + VA^{-1}U \right)^{-1} VA^{-1}$$

shown that:

$$\lim_{k\to+\infty} Q_{k,i}^{-1} = \lim_{k\to+\infty} \left(\frac{1}{\sigma_i^2} I_k - \frac{1}{k\sigma_i^2} S_{k,i} S_{k,i}^H + \frac{1}{k^2} S_{k,i} \left(\Phi_i + \frac{\sigma_i^2}{k} \right)^{-1} S_{k,i}^H \right) \quad (12)$$

On the other hand, let us express $R_{k,i}$ from $Q_{k+1,i}$ by considering that $\mathbb{0}_{i,j}$ is a $i \times j$ matrix of zeros:

$$\begin{aligned} R_{k,i} &= \begin{bmatrix} I_k & \mathbb{0}_{k,1} \end{bmatrix} Q_{k+1,i} \begin{bmatrix} \mathbb{0}_{k,1} \\ 1 \end{bmatrix} \\ &\underset{(2),(6)}{=} \begin{bmatrix} I_k & \mathbb{0}_{k,1} \end{bmatrix} \left(S_{k+1,i} \Phi_i S_{k+1,i}^H + \sigma_i^2 I_{k+1} \right) \begin{bmatrix} \mathbb{0}_{k,1} \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} I_k & \mathbb{0}_{k,1} \end{bmatrix} S_{k+1,i} \Phi_i S_{k+1,i}^H \begin{bmatrix} \mathbb{0}_{k,1} \\ 1 \end{bmatrix} \\ &= S_{k,i} \Phi_i \begin{bmatrix} e^{jk\theta_{1,i}} & \dots & e^{jk\theta_{M_i,i}} \end{bmatrix}^H = S_{k,i} \Phi_i \Psi_{k,i}^H \end{aligned} \quad (13)$$

where $\Psi_{k,i} = \begin{bmatrix} e^{jk\theta_{1,i}} & \dots & e^{jk\theta_{M_i,i}} \end{bmatrix}$.
Consequently, using (12) and (13), one can express $\lim_{k\to+\infty} R_{k,i}^H Q_{k,i}^{-1} R_{k,i}$ as the sum of three terms given below:

1. The first one is defined as:

$$\frac{1}{\sigma_i^2} \lim_{k\to+\infty} R_{k,i}^H R_{k,i} = \frac{1}{\sigma_i^2} \lim_{k\to+\infty} \Psi_{k,i} \Phi_i S_{k,i}^H S_{k,i} \Phi_i \Psi_{k,i}^H \quad (14)$$

   Due to the asymptotic orthogonal properties (5) between the columns of the matrix $S_{k,i}$, $S_{k,i}^H S_{k,i}$ can be replaced by kI_k when k increases and tends to infinity and hence one gets:

$$\frac{1}{\sigma_i^2} \lim_{k\to+\infty} R_{k,i}^H R_{k,i} = \lim_{k\to+\infty} \frac{k}{\sigma_i^2} \Psi_{k,i} \Phi_i^2 \Psi_{k,i}^H \quad (15)$$

2. Concerning the second term, one has:

$$\begin{aligned} \lim_{k\to+\infty} -\frac{1}{k\sigma_i^2} R_{k,i}^H S_{k,i} S_{k,i}^H R_{k,i} &= \lim_{k\to+\infty} -\frac{1}{k\sigma_i^2} \Psi_{k,i} \Phi_i S_{k,i}^H S_{k,i} S_{k,i}^H S_{k,i} \Phi_i \Psi_{k,i}^H \\ &= \lim_{k\to+\infty} -\frac{k}{\sigma_i^2} \Psi_{k,i} \Phi_i^2 \Psi_{k,i}^H \end{aligned} \quad (16)$$

3. Finally, the third term satisfies:

$$\begin{aligned}
&\lim_{k\to+\infty} \frac{1}{k^2} R_{k,i}^H S_{k,i} (\Phi_i + \frac{\sigma_i^2}{k})^{-1} S_{k,i}^H R_{k,i} \qquad (17)\\
&= \lim_{k\to+\infty} \frac{1}{k^2} \Psi_{k,i} \Phi_i S_{k,i}^H S_{k,i} (\Phi_i + \frac{\sigma_i^2}{k})^{-1} S_{k,i}^H S_{k,i} \Phi_i \Psi_{k,i}^H \\
&= \lim_{k\to+\infty} \Psi_{k,i} \Phi_i \Psi_{k,i}^H = Tr(\Phi_i)
\end{aligned}$$

where $Tr(A)$ is the trace of the matrix A.

Consequently, given (12), (15), (16) and (17), the limit of $R_{k,i}^H Q_{k,i}^{-1} R_{k,i}$ when k increases and tends to infinity can be rewritten as:

$$\lim_{k\to+\infty} R_{k,i}^H Q_{k,i}^{-1} R_{k,i} = Tr(\Phi_i) \qquad (18)$$

Since $r_{0,i} = Tr(\Phi_i) + \sigma_i^2$, one can deduce that:

$$\lim_{k\to+\infty} \frac{|Q_{k+1,i}|}{|Q_{k,i}|} = r_{0,i} - Tr(\Phi_i) = Tr(\Phi_i) + \sigma_i^2 - Tr(\Phi_i) = \sigma_i^2 \qquad (19)$$

Remark.

- *When k is larger than the number M_i of complex exponentials, there are M_i predominant eigenvalues defining the so-called signal subspace and $k - M_i$ eigenvalues equal to the variance of the additive noise, the associated eigenvectors spanning the noise subspace. Thus, when the process is a complex exponential disturbed by an additive white noise, $Q_{k,i}$ has a single predominant eigenvalue equal to $k\gamma_{1,i} + \sigma_i^2$ and $k-1$ eigenvalues equal to σ_i^2. Therefore $|Q_{k,i}| = (k\gamma_{1,i} + \sigma_i^2) \times \sigma_i^{2(k-1)}$ and one has:*

$$\lim_{k\to+\infty} \frac{|Q_{k+1,i}|}{|Q_{k,i}|} = \lim_{k\to+\infty} \frac{((k+1)\gamma_{1,i} + \sigma_i^2)\sigma_i^{2k}}{(k\gamma_{1,i} + \sigma_i^2)\sigma_i^{2(k-1)}} = \sigma_i^2 \qquad (20)$$

- *The weighted sum of the two covariance matrices $Q_{k,1}$ and $Q_{k,2}$ leads to a covariance matrix[6] $Q_{wei,k}(b,a) = bQ_{k,1} + aQ_{k,2}$, with $b > 0$ and $a > 0$ associated with a third process which is also a weighted sum of complex exponentials disturbed by an additive noise. If the two processes do not share*

[6] Subscript $_{wei}$ refers to the fact that the matrix is a weighted sum of the covariance matrices of the two processes.

a common normalized angular frequency, the number of complex exponentials is equal to M_1+M_2. The variances $\{\gamma_{m,1}\}_{m=1,\ldots,M_1}$ are multiplied by b while $\{\gamma_{m,2}\}_{m=1,\ldots,M_2}$ are weighted by a. When they have one common normalized angular frequency, the number of complex exponentials reduces to M_1+M_2-1. When they have two common normalized angular frequencies, the number reduces to M_1+M_2-2 and so on. The variance of the additive noise is equal $b\sigma_1^2+a\sigma_2^2$.

When comparing two processes, different approaches can be considered. The model parameters and/or the spectrum can be compared. An alternative consists in using divergences. For this reason, in the next section, we propose a brief state of the art on the divergence that will be addressed in this study.

3. Brief state of the art of the divergences addressed in this study

3.1. Definitions

To study the dissimilarities between pdfs, divergences can be used. In Figure 1 below, we propose a unified view of a wide class of divergences between two pdfs denoted p_1 and p_2. The most popular one is the KL divergence which satisfies:

$$KL_k^{(1,2)} = \int_{X_k} p_1(X_k) \ln\left(\frac{p_1(X_k)}{p_2(X_k)}\right) \mathrm{d}X_k \tag{21}$$

Alternative divergences can be expressed as a function g of the Chernoff coefficient C_α of order α that is computed between p_1 and p_2, which is defined by:

$$C_\alpha\big(p_1(X_k), p_2(X_k)\big) = \int_{X_k} p_1^\alpha(X_k) p_2^{1-\alpha}(X_k) dX_k \tag{22}$$

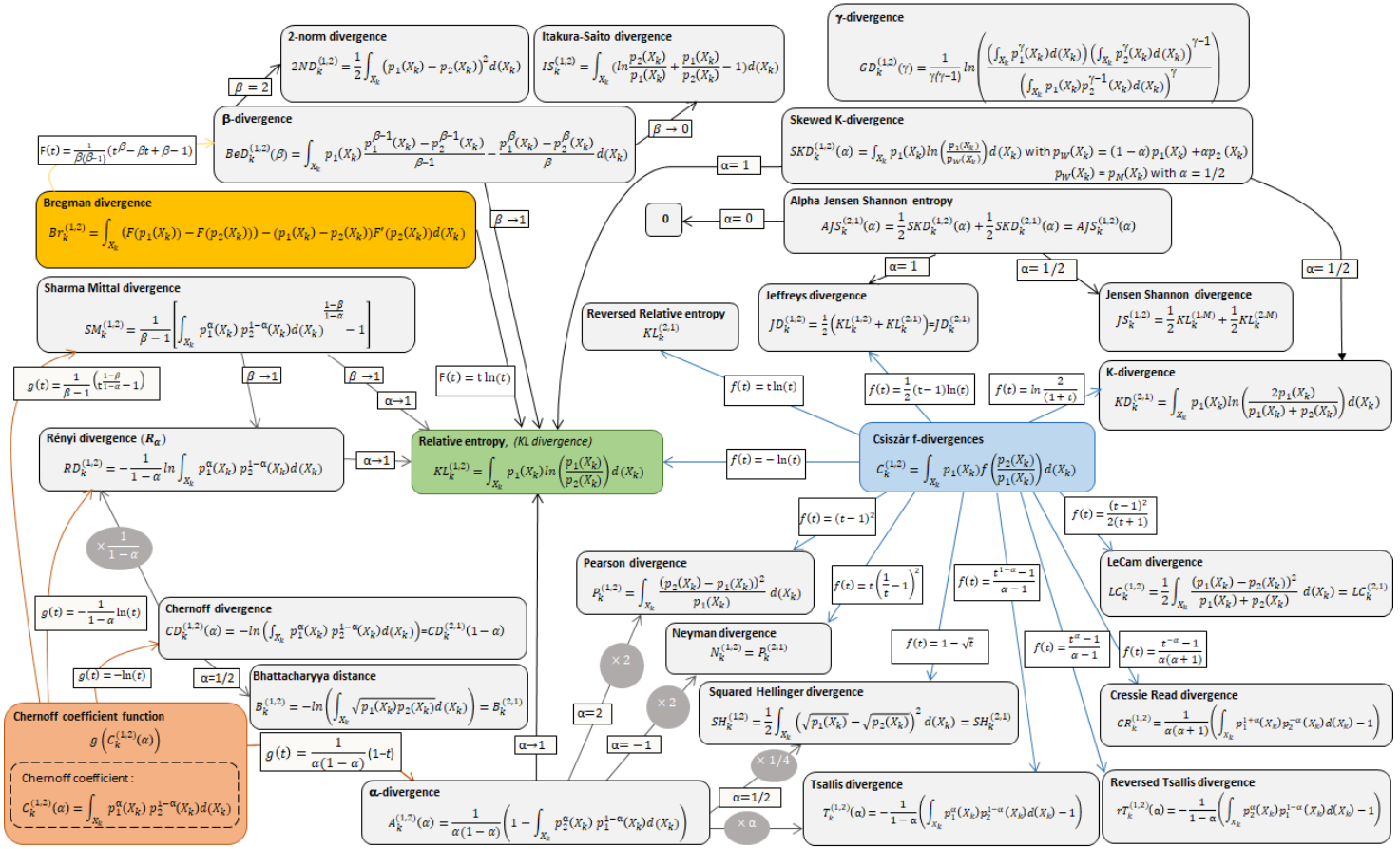

Figure 1: Links between different divergences

Remark. *It corresponds to the integral* $I_1(p_1, p_2, a, b) = \int_{X_k} p_1^a(X_k) p_2^b(X_k) dX_k$ *computed for* $a = \alpha$ *and* $b = 1 - \alpha$ *with* $0 < \alpha < 1$.

Using the above definition, it can be shown that the Chernoff coefficient is self dual, meaning that:

$$C_\alpha\big(p_1(X_k), p_2(X_k)\big) = C_{1-\alpha}\big(p_2(X_k), p_1(X_k)\big) \tag{23}$$

When α is equal to 0 or 1, the Chernoff coefficient is equal to 1 due to the property of the pdfs that sum to 1.

Let us now give some examples of divergences that are functions of the Chernoff coefficient. When the function $g(t) = -\ln(t)$, the Chernoff divergence between p_1 and p_2, denoted as $CD_k^{(1,2)}(\alpha)$, is defined as follows [5]:

$$CD_k^{(1,2)}(\alpha) = -\ln\Big(C_\alpha(p_1(X_k), p_2(X_k))\Big) \tag{24}$$

Given the remark above, the Chernoff divergence is necessarily null when α is equal to 0 or 1. By setting α to $\frac{1}{2}$, the Bhattacharyya distance denoted as $BD_k^{(1,2)}$ can be deduced [3] and has the advantage of being symmetric:

$$\begin{aligned} BD_k^{(1,2)} &= CD_k^{(1,2)}(\frac{1}{2}) = -\ln\Big(C_{\frac{1}{2}}(p_1(X_k), p_2(X_k))\Big) \\ &= -\ln\Big(\int_{X_k} \sqrt{p_1(X_k) p_2(X_k)} dX_k\Big) = BD_k^{(2,1)} \end{aligned} \tag{25}$$

The Battacharyya distance depends on the so-called "fidelity" equal to $\int_{X_k} \sqrt{p_1(X_k) p_2(X_k)} dX_k$. The latter is in the interval $[0, 1]$ and equal to unity only if the two pdfs are identical. Hence, $BD_k^{(1,2)} \geq 0$.

When $g(t) = -\frac{1}{1-\alpha}\ln(t)$, the Rényi divergence of order α [39] can be obtained and corresponds to the Chernoff divergence up to a multiplicative factor equal to $\frac{1}{1-\alpha}$:

$$RD_k^{(1,2)}(\alpha) = -\frac{1}{1-\alpha}\ln\int_{X_k} p_1^\alpha(X_k)\, p_2^{1-\alpha}(X_k)\, \mathrm{d}X_k = \frac{1}{1-\alpha} CD_k^{(1,2)}(\alpha) \tag{26}$$

By using L'Hôspital rule, the Rényi divergence is shown to tend to the KL divergence when α tends to 1. Rényi divergence for several univariate and multivariate distributions can be found in the paper written by Gil *et al.* [14].

In addition, the equalities in (26) can be rewritten this way:

$$C_\alpha\big(p_1(X_k), p_2(X_k)\big) = \exp\big(-CD_k^{(1,2)}(\alpha)\big) = \exp\big((\alpha - 1) RD_k^{(1,2)}(\alpha)\big) \tag{27}$$

The α-divergence, denoted in this paper as $A_k^{(1,2)}(\alpha)$, can be also considered and derived from the Chernoff coefficient. Different definitions exist [1]:

$$A_k^{(1,2)}(\alpha) = -\frac{1}{\alpha(1-\alpha)} \int_{X_k} p_1^{\alpha}(X_k)\, p_2^{1-\alpha}(X_k) - \alpha p_1(X_k) + (\alpha-1) p_2(X_k)\, \mathrm{d}X_k \quad (28)$$

As the integrals of the pdfs are equal to 1, the above expression can be simplified as follows:

$$\begin{aligned} A_k^{(1,2)}(\alpha) &= -\frac{1}{\alpha(1-\alpha)} \Big(\int_{X_k} p_1^{\alpha}(X_k)\, p_2^{1-\alpha}(X_k)\, \mathrm{d}X_k - 1 \Big) \\ &= -\frac{1}{\alpha(1-\alpha)} \Big(C_\alpha\big(p_1(X_k), p_2(X_k)\big) - 1 \Big) \end{aligned} \quad (29)$$

This means that $g(t) = -\frac{1}{\alpha(1-\alpha)}(t-1)$ is used to retrieve the α-divergence from the Chernoff coefficient. Note that by combining (27) and (29), this leads to:

$$\begin{aligned} A_k^{(1,2)}(\alpha) &= -\frac{1}{\alpha(1-\alpha)} \Big(\exp\big(-CD_k^{(1,2)}(\alpha)\big) - 1 \Big) \\ &= -\frac{1}{\alpha(1-\alpha)} \Big(\exp\big((\alpha-1) RD_k^{(1,2)}(\alpha)\big) - 1 \Big) \end{aligned} \quad (30)$$

Some other comments can be made at this stage:

1. The α-divergence is also self-dual because $A_k^{(1,2)}(\alpha) = A_k^{(2,1)}(1-\alpha)$.
2. One can use the Amari notation, *i.e.* instead of defining the α-divergence from the variate $\alpha = \frac{(1-\alpha_a)}{2}$, α_a is used. When α takes values between 0 and 1, the value taken by α_a varies from -1 to 1. In addition, using the Amari notation, $A_k^{(1,2)}(\alpha_a) = A_k^{(2,1)}(-\alpha_a)$.
3. Using the L'Hôspital rule on (28), it can be shown that the α divergence tends to the KL between p_1 and p_2 when α tends to 1 and to the KL between p_2 and p_1 when α tends to 0.
4. One can retrieve three other divergences, namely the Pearson divergence $P_k^{(1,2)}$, the Neyman divergence $N_k^{(1,2)}$ and the squared Hellinger distance $SH_k^{(1,2)}$ by setting α to -1, 2 and $\frac{1}{2}$ respectively:

$$\begin{aligned} P_k^{(1,2)} = 2A_k^{(1,2)}(-1) &= \int_{X_k} \frac{(p_2(X_k) - p_1(X_k))^2}{p_1(X_k)} \mathrm{d}X_k \\ &= \int_{X_k} p_2^2(X_k)\, p_1^{-1}(X_k)\, \mathrm{d}X_k - 1 \end{aligned} \quad (31)$$

$$N_k^{(1,2)} = 2A_k^{(1,2)}(2) = P_k^{(2,1)} \tag{32}$$

$$\begin{aligned} SH_k^{(1,2)} = \frac{1}{4}A_k^{(1,2)}(\frac{1}{2}) &= \frac{1}{2}\int_{X_k}\Big(\sqrt{(p_1\,(X_k)} - \sqrt{p_2\,(X_k)}\Big)^2 \mathrm{d}X_k \\ &= 1 - \int_{X_k}\sqrt{p_1\,(X_k)\,p_2\,(X_k)}\mathrm{d}X_k = SH_k^{(2,1)} \\ &= 1 - \exp\big(-BD_k^{(1,2)}\big) \end{aligned} \tag{33}$$

Note that the squared Hellinger distance $SH_k^{(1,2)}$ also depends on the fidelity. It is symmetric and in the range $[0, 1]$. Sometimes, the normalization factor $\frac{1}{2}$ is omitted.

5. The Tsallis divergence $T_k^{(1,2)}(\alpha)$, the reversed Tsallis divergence $rT_k^{(1,2)}(\alpha)$ and the Cressie-Read divergence $CR_k^{(1,2)}(\alpha)$ are closely related to the above family. Indeed given (29), one has:

$$\begin{aligned} T_k^{(1,2)}(\alpha) &= -\frac{1}{(1-\alpha)}\Big(C_\alpha\big(p_1(X_k), p_2(X_k)\big) - 1\Big) = \alpha A_k^{(1,2)}(\alpha) \\ &= -\frac{1}{(1-\alpha)}\Big(\exp\big((\alpha-1)RD_k^{(1,2)}(\alpha)\big) - 1\Big) \end{aligned} \tag{34}$$

$$\begin{aligned} rT_k^{(1,2)}(\alpha) &= -\frac{1}{(1-\alpha)}\Big(C_\alpha\big(p_2(X_k), p_1(X_k)\big) - 1\Big) \\ &= -\frac{1}{(1-\alpha)}\Big(C_{1-\alpha}\big(p_1(X_k), p_2(X_k)\big) - 1\Big) = \alpha A_k^{(1,2)}(1-\alpha) \end{aligned} \tag{35}$$

$$\begin{aligned} CR_k^{(1,2)}(\alpha) &= \frac{1}{\alpha(\alpha+1)}\Big(\int_{X_k} p_1^{1+\alpha}\,(X_k)\,p_2^{-\alpha}\,(X_k)\,\mathrm{d}X_k - 1\Big) \\ &= A_k^{(1,2)}(\alpha+1) = \frac{T_k^{(1,2)}(\alpha+1)}{\alpha+1} \end{aligned} \tag{36}$$

Remark. *The Pearson divergence, the Neyman divergence, the squared Hellinger distance, the reversed Tsallis divergence and the Cressie-Read divergence are special cases of Csiszar divergence also known as Csiszar-Morimoto or Ali-Silvey divergence. Indeed, it is defined by:*

$$D_f\big(p_1(X_k), p_2(X_k)\big) = \int_{X_k} p_1(X_k) f\Big(\frac{p_2(X_k)}{p_1(X_k)}\Big) dX_k \tag{37}$$

where the function $f(t)$ is convex and $f(1) = 0$. One can retrieve the expressions of the divergence by respectively choosing the convex function f equal to $(t-1)^2$, $t(\frac{1}{t}-1)^2$, $1-\sqrt{t}$, $\frac{t^{1-\alpha}-1}{\alpha-1}$ and $\frac{t^{-\alpha}-1}{\alpha(\alpha+1)}$. Moreover, the Chernoff coefficient can be expressed as $D_f\big(p_1(X_k), p_2(X_k)\big) + 1$, where $f(t) = t^{1-\alpha} - 1$.

Finally, the SM divergence of parameters β and α between p_1 and p_2 is given by:

$$SM_k^{(1,2)}(\beta,\alpha) = \frac{1}{\beta-1}\Big(C_\alpha^{\frac{1-\beta}{1-\alpha}}(p_1(X_k), p_2(X_k)) - 1\Big) \tag{38}$$

When $\beta = \alpha$, the SM divergence coincides with the Tsallis divergence:

$$SM_k^{(1,2)}(\alpha,\alpha) = \frac{1}{\alpha-1}\Big(C_\alpha(p_1(X_k), p_2(X_k)) - 1\Big) = \frac{1}{\alpha}A_k^{(1,2)}(\alpha) = T_k^{(1,2)}(\alpha) \tag{39}$$

In addition, using (27) and (38), one can see that:

$$SM_k^{(1,2)}(\beta,\alpha) = \frac{1}{\beta-1}\Big(\exp\Big[(\beta-1)RD_k^{(1,2)}(\alpha)\Big] - 1\Big) \tag{40}$$

The SM divergence encompasses various divergences. When β tends to 1, $SM_k^{(1,2)}(\beta,\alpha)$ tends to $RD_k^{(1,2)}(\alpha)$. So, when β tends to 1 and α tends to 1, $SM_k^{(1,2)}(\beta,\alpha)$ tends to the KL divergence.

One can easily define symmetric versions of these divergences by taking the minimum value, the sum or the mean of the divergence computed between p_1 and p_2 and the one between p_2 and p_1. This is for instance the case when dealing with the Jeffreys divergence [21] which is the symmetric version of the KL divergence. An alternative is to sum the divergence between the first pdf and the pdf mean and the divergence between the second pdf and the pdf mean. Thus, the Jensen–Shannon divergence (JSD) is the illustration of this alternative to get a symmetric version of the KL divergence [26].

Two other families of divergence can be considered. One the one hand, the β-divergence denoted as $BeD_k^{(1,2)}(\beta)$ can be expressed as follows:

$$BeD_k^{(1,2)}(\beta) = \int_{X_k} p_1(X_k)\frac{p_1^{\beta-1}(X_k) - p_2^{\beta-1}(X_k)}{\beta-1} - \frac{p_1^{\beta}(X_k) - p_2^{\beta}(X_k)}{\beta} dX_k \tag{41}$$

$$\begin{aligned} &= \frac{1}{\beta(\beta-1)}\int_{X_k} p_1^{\beta}(X_k)dX_k + \frac{1}{\beta}\int_{X_k} p_2^{\beta}(X_k)dX_k \\ &- \frac{1}{\beta-1}\int_{X_k} p_1(X_k)p_2^{\beta-1}(X_k)dX_k \end{aligned}$$

By introducing the function[7] $\ln_\beta(t) = \frac{t^{1-\beta}-1}{1-\beta}$, the above expression (41) can be rewritten as follows:

$$BeD_k^{(1,2)}(\beta) = -\frac{1}{\beta}\int_{X_k}\left(p_1^{\beta}(X_k)\ln_{\frac{1}{\beta}}\frac{p_2^{\beta}(X_k)}{p_1^{\beta}(X_k)}\right) + p_1^{\beta}(X_k) - p_2^{\beta}(X_k)dX_k \tag{42}$$

If $I_2(p_1,\beta) = \int_{X_k} p_1^{\beta}(X_k)dX_k$ and $I_1(p_1,p_2,1,\beta) = \int_{X_k} p_1(X_k)p_2^{\beta}(X_k)dX_k$ are considered, (41) becomes:

$$BeD_k^{(1,2)}(\beta) = \frac{1}{\beta(\beta-1)}I_2(p_1,\beta) + \frac{1}{\beta}I_2(p_2,\beta) - \frac{1}{\beta-1}I_1(p_1,p_2,1,\beta-1) \tag{43}$$

Remark. *It should be noted the β-divergence is a special case of the Bregman divergence, as shown on Figure 1.*

Among the particular cases, one can notice that when β tends to 0, the β-divergence reduces to the Itakura-Saito divergence. Indeed, given (41) and applying the L'Hôspital rule to the second term, one obtains:

$$\lim_{\beta\to 0} BeD_k^{(1,2)}(\beta) = \int_{X_k}\Big(\ln\big(\frac{p_2(X_k)}{p_1(X_k)}\big) + \frac{p_1(X_k)}{p_2(X_k)} - 1\Big)dX_k = IS_k^{(1,2)} \tag{44}$$

If β tends to 1, and considering (42), one gets:

$$\begin{aligned} \lim_{\beta\to 1} BeD_k^{(1,2)}(\beta) &= -\int_{X_k} p_1(X_k)\ln\frac{p_2(X_k)}{p_1(X_k)} + p_1(X_k) - p_2(X_k)dX_k \\ &= \int_{X_k} p_1(X_k)\ln\frac{p_1(X_k)}{p_2(X_k)}dX_k \end{aligned} \tag{45}$$

[7] when β tends 1, $\ln_\beta(t)$ tends to $\ln(t)$

One hence retrieves the KL divergence.

If $\beta = 2$, (41) becomes:

$$BeD_k^{(1,2)}(2) = \frac{1}{2}\int_{X_k}\Big(p_1(X_k) - p_2(X_k)\Big)^2 dX_k = BeD_k^{(2,1)}(2) \tag{46}$$

which can be seen as the 2-norm of the pdf difference up to a multiplicative factor equal to $\frac{1}{2}$. Note that it is symmetric.

Finally, let us introduce the γ-divergence [11] denoted as $G_k^{(1,2)}(\gamma)$ defined as :

$$\begin{aligned} G_k^{(1,2)}(\gamma) &= \frac{1}{\gamma(\gamma-1)}\ln\left(\frac{\left(\int_{X_k} p_1^\gamma(X_k)dX_k\right)\left(\int_{X_k} p_2^\gamma(X_k)dX_k\right)^{\gamma-1}}{\left(\int_{X_k} p_1(X_k)p_2^{\gamma-1}(X_k)dX_k\right)^{\gamma}}\right) \\ &= \frac{1}{\gamma(\gamma-1)}\ln\left(\frac{I_2(p_1,\gamma)I_2^{\gamma-1}(p_2,\gamma)}{I_1^\gamma(p_1,p_2,1,\gamma-1)}\right) \end{aligned} \tag{47}$$

and the $\alpha\gamma$-divergence denoted as $AG_k^{(1,2)}(\gamma)$ which is deduced from the definition (28) of the α divergence by considering the variable γ instead of α and the mapping $\alpha\int_{X_k} p_1^\beta(X_k)p_1^\gamma(X_k)dX_k \to \alpha\ln\Big(\int_{X_k} p_1^\beta(X_k)p_1^\gamma(X_k)dX_k\Big)$:

$$\begin{aligned} AG_k^{(1,2)}(\gamma) &= \frac{1}{\gamma(\gamma-1)}\ln\left(\frac{\int_{X_k} p_1^\gamma(X_k)p_2^{1-\gamma}(X_k)dX_k}{\left(\int_{X_k} p_1(X_k)dX_k\right)^{\gamma}\left(\int_{X_k} p_2(X_k)dX_k\right)^{1-\gamma}}\right) \\ &= \frac{1}{\gamma(\gamma-1)}\ln\left(\int_{X_k} p_1^\gamma(X_k)p_2^{1-\gamma}(X_k)dX_k\right) = \frac{1}{\gamma}RD_k^{(1,2)}(\gamma) \end{aligned} \tag{48}$$

since p_1 and p_2 are pdfs whose integrals are equal to 1. It is hence proportional to the Rényi divergence of order α by choosing $\alpha = \gamma$. Therefore, when γ tends to 1, the $\alpha\gamma$-divergence tends to the KL divergence.

Given the definitions of these various divergences, let us now look at their expressions when dealing with the Gaussian case.

3.2. Expressions of the divergences in the Gaussian case

Let us introduce the pdfs related to two real Gaussian random vectors defined from k consecutive samples, *i.e.* $X_{k,i} = [x_{t,i}\ x_{t-1,i}\ \cdots\ x_{t-k+1,i}]^T$ for $i = 1, 2$:

$$p_i(X_{k,i}) = \frac{1}{(\sqrt{2\pi})^k|Q_{k,i}|^{1/2}}\exp\big(-\frac{1}{2}[X_{k,i}-\mu_{k,i}]^T Q_{k,i}^{-1}[X_{k,i}-\mu_{k,i}]\big) \tag{49}$$

with $\mu_{k,i} = E[X_{k,i}]$ the mean, $|Q_{k,i}|$ the determinant of the covariance matrix $Q_{k,i} = E[(X_{k,i} - \mu_{k,i})(X_{k,i} - \mu_{k,i})^T]$ and $E[\cdot]$ the expectation. Given[8]

$$Q_{k,wei}(1-\alpha,\alpha) = (1-\alpha)Q_{k,1} + \alpha Q_{k,2} \tag{50}$$

and

$$\Delta\mu_k = \mu_{k,2} - \mu_{k,1} \tag{51}$$

we can show, by substituting the pdfs by their expressions in the definition (24) and after some mathematical developments (See Appendix A.2 and equation (A.20) for proof), that the Chernoff coefficient is given by:

$$\begin{aligned} C_\alpha\big(p_1(X_k), p_2(X_k)\big) = & \frac{|Q_{k,1}|^{\frac{1-\alpha}{2}}|Q_{k,2}|^{\frac{\alpha}{2}}}{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{1}{2}}} \\ & \times \exp\left(-\frac{\alpha(1-\alpha)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\right) \end{aligned} \tag{52}$$

This result can be obtained differently by using the approach proposed by Nielsen in [31] and recalled in Appendix B. It should be noted that $Q_{k,wei}(1-\alpha,\alpha)$ must be definite positive. Therefore, this result is only valid[9] when $\alpha \in]0,1[$. Consequently the Chernoff divergence is the following:

$$CD_k^{(1,2)}(\alpha) = \frac{1}{2}\ln\left(\frac{|Q_{k,wei}(1-\alpha,\alpha)|}{|Q_{k,1}|^{1-\alpha}|Q_{k,2}|^{\alpha}}\right) + \frac{\alpha(1-\alpha)}{2}Tr(Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Delta\mu_k^T) \tag{53}$$

The Bhattacharyya distance can be deduced as follows:

$$BD_k^{(1,2)} = CD_k^{(1,2)}(\frac{1}{2}) = \frac{1}{2}\ln\left(\frac{|Q_{k,wei}(\frac{1}{2},\frac{1}{2})|}{|Q_{k,1}|^{\frac{1}{2}}|Q_{k,2}|^{\frac{1}{2}}}\right) + \frac{1}{8}Tr(Q_{k,wei}^{-1}(\frac{1}{2},\frac{1}{2})\Delta\mu_k\Delta\mu_k^T) \tag{54}$$

Concerning the expression of the Rényi divergence, given (26) which gives the link with Chernoff divergence and provided that α is in the interval $]0,1[$ to

[8] The definition of $Q_{k,wei}(1-\alpha,\alpha)$ amounts to saying that a third process is introduced and corresponds to a linear combination of the two processes to be compared.

[9] For information, the KL divergence satisfies in the Gaussian case [14]:

$$KL_k^{(1,2)} = \frac{1}{2}\Big[Tr(Q_{k,2}^{-1}Q_{k,1}) - k - \ln\frac{|Q_{k,1}|}{|Q_{k,2}|} + Tr(Q_{2,k}^{-1}\Delta\mu_k\Delta\mu_k^T)\Big]$$

ensure that $Q_{k,wei}(1-\alpha,\alpha)$ is not definite non-negative, this leads to:

$$RD_k^{(1,2)}(\alpha) = \frac{1}{2(1-\alpha)} \ln\Big(\frac{|Q_{k,wei}(1-\alpha,\alpha)|}{|Q_{k,1}|^{1-\alpha}|Q_{k,2}|^{\alpha}}\Big) + \frac{\alpha}{2} Tr(Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Delta\mu_k^T) \tag{55}$$

There are two consequences. On the one hand, the expression of the $\alpha\gamma$-divergence can be deduced using (48). Provided that γ is in the interval $]0,1[$ to ensure that $Q_{k,wei}(1-\gamma,\gamma)$ is not definite non-negative, this leads to:

$$AG_k^{(1,2)}(\gamma) = \frac{1}{2(1-\gamma)\gamma} \ln\Big(\frac{|Q_{k,wei}(1-\gamma,\gamma)|}{|Q_{k,1}|^{1-\gamma}|Q_{k,2}|^{\gamma}}\Big) + \frac{1}{2} Tr(Q_{k,wei}^{-1}(1-\gamma,\gamma)\Delta\mu_k\Delta\mu_k^T) \tag{56}$$

On the other hand, combining (29) and (52), the expression of the α-divergence in the Gaussian case is:

$$\begin{aligned} &A_k^{(1,2)}(\alpha) \qquad (57)\\ &= \frac{1}{\alpha(1-\alpha)}\Big(1 - \frac{|Q_{k,1}|^{\frac{1-\alpha}{2}}|Q_{k,2}|^{\frac{\alpha}{2}}}{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{1}{2}}} \exp\Big[\frac{\alpha(\alpha-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Big]\Big) \end{aligned}$$

Therefore, using (33), (34) and (35), one can deduce the expression of the squared Hellinger distance, the Tsallis and reversed Tsallis divergences, but we cannot obtain the Pearson and Neyman divergences and the Cressie-Read from the expression above since the values of α are not in the interval $]0,1[$:

$$SH_k^{(1,2)} = 1 - \frac{|Q_{k,1}|^{\frac{1}{4}}|Q_{k,2}|^{\frac{1}{4}}}{|Q_{k,wei}(\frac{1}{2},\frac{1}{2})|^{\frac{1}{2}}} \exp\Big[-\frac{1}{8}\Delta\mu_k^T Q_{k,wei}^{-1}(\frac{1}{2},\frac{1}{2})\Delta\mu_k\Big)\Big] \tag{58}$$

$$\begin{aligned} &T_k^{(1,2)}(\alpha) \qquad (59)\\ &= \frac{1}{1-\alpha}\Big(1 - \frac{|Q_{k,1}|^{\frac{1-\alpha}{2}}|Q_{k,2}|^{\frac{\alpha}{2}}}{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{1}{2}}} \exp\Big[\frac{\alpha(\alpha-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Big]\Big) \end{aligned}$$

$$\begin{aligned} &rT_k^{(1,2)}(\alpha) \qquad (60)\\ &= \frac{1}{1-\alpha}\Big(1 - \frac{|Q_{k,1}|^{\frac{\alpha}{2}}|Q_{k,2}|^{\frac{1-\alpha}{2}}}{|Q_{k,wei}(\alpha,1-\alpha)|^{\frac{1}{2}}} \exp\Big[\frac{\alpha(\alpha-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\alpha,1-\alpha)\Delta\mu_k\Big]\Big) \end{aligned}$$

Given (40), the SM divergence becomes:

$$\begin{aligned} &SM_k^{(1,2)}(\beta,\alpha) \qquad (61)\\ &= \frac{1}{\beta-1}\Big(\frac{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{\beta-1}{2(1-\alpha)}}}{|Q_{k,1}|^{\frac{\beta-1}{2}}|Q_{k,2}|^{\frac{\alpha(\beta-1)}{2(1-\alpha)}}} \exp\Big[\frac{\alpha(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Big] - 1\Big) \end{aligned}$$

For the β and γ-divergences, we first propose to use two results we have detailed in Appendix A and that deal with the analytical expressions of $I_2(p_1,\beta) = \int_{X_k} p_1^\beta(X_k)dX_k$ and $I_1(p_1,p_2,1,\beta) = \int_{X_k} p_1(X_k)p_2^\beta(X_k)dX_k$ in the Gaussian case. Thus, given (41), (A.2) and (A.19) one has:

$$BeD_k^{(1,2)}(\beta) = \frac{1}{\beta^{\frac{k}{2}+1}(2\pi)^{\frac{k(\beta-1)}{2}}}\left(\frac{1}{(\beta-1)|Q_{k,1}|^{\frac{(\beta-1)}{2}}} + \frac{1}{|Q_{k,2}|^{\frac{(\beta-1)}{2}}}\right) \tag{62}$$
$$- \frac{1}{\beta-1}\frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}\exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)$$

where $Q_{k,wei}(\beta-1,1) = (\beta-1)Q_{k,1} + Q_{k,2}$ (See Appendix A.2).

In the following, only values of β larger than 1 will be considered so that $Q_{k,wei}(\beta-1,1)$ can be definite positive without any doubt. Thus, when $\beta = 2$, one gets:

$$BeD_k^{(1,2)}(2) = \frac{1}{(2\pi)^{\frac{k}{2}}}\Big[\left(\frac{|Q_{k,1}|^{\frac{1}{2}} + |Q_{k,2}|^{\frac{1}{2}}}{2^{\frac{k}{2}+1}|Q_{k,1}|^{\frac{1}{2}}|Q_{k,2}|^{\frac{1}{2}}}\right) \tag{63}$$
$$- \frac{1}{|Q_{k,wei}(1,1)|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(1,1)\Delta\mu_k\right)\Big]$$

In addition, starting from (47) and using (A.2) and (A.19), one gets:

$$G_k^{(1,2)}(\gamma) = \ln\left(\frac{|Q_{k,wei}(\gamma-1,1)|^{\frac{1}{2(\gamma-1)}}}{|Q_{k,1}|^{\frac{1}{2\gamma}}|Q_{k,2}|^{\frac{1}{2\gamma(\gamma-1)}}}\right) - \frac{k\ln\gamma}{2(\gamma-1)} \tag{64}$$
$$+ \frac{1}{2}\mathrm{Tr}\left(Q_{k,wei}^{-1}(\gamma-1,1)\Delta\mu_k\Delta\mu_k^T\right)$$

where $Q_{k,wei}(\gamma-1,1) = (\gamma-1)Q_{k,1} + Q_{k,2}$ (See Appendix A.2).

In the following, let us analyze how the divergences evolve when k increases when dealing with two NSCE processes.

4. Analysis of the increments of the divergences

4.1. Expression of the increments

Taking into account (53), (54) and (55), let us first express the increments $\Delta CD_k^{(1,2)}(\alpha) = CD_{k+1}^{(1,2)}(\alpha) - CD_k^{(1,2)}(\alpha)$, $\Delta BD_k^{(1,2)}$and $\Delta RD_k^{(1,2)}(\alpha)$ which are similarly defined. Given (53), one has:

$$\Delta CD_k^{(1,2)}(\alpha) = \frac{1}{2}\ln\Big(\frac{|Q_{k,2}|^\alpha}{|Q_{k+1,2}|^\alpha}\frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|}\frac{|Q_{k,1}|^{1-\alpha}}{|Q_{k+1,1}|^{1-\alpha}}\Big) \tag{65}$$
$$+ \frac{\alpha(1-\alpha)}{2}\Big(\mathrm{Tr}(Q_{k+1,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_{k+1}\Delta\mu_{k+1}^T) - \mathrm{Tr}(Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Delta\mu_k^T)\Big)$$

$$\begin{aligned}\Delta BD_k^{(1,2)} &= \frac{1}{2}\ln\Big(\frac{|Q_{k,2}|^{\frac{1}{2}}}{|Q_{k+1,2}|^{\frac{1}{2}}}\frac{|Q_{k+1,wei}(\frac{1}{2},\frac{1}{2})|}{|Q_{k,wei}(\frac{1}{2},\frac{1}{2})|}\frac{|Q_{k,1}|^{\frac{1}{2}}}{|Q_{k+1,1}|^{\frac{1}{2}}}\Big) \\ &+ \frac{1}{8}\Big(\mathrm{Tr}(Q_{k+1,wei}^{-1}(\frac{1}{2},\frac{1}{2})\Delta\mu_{k+1}\Delta\mu_{k+1}^T) - \mathrm{Tr}(Q_{k,wei}^{-1}(\frac{1}{2},\frac{1}{2})\Delta\mu_k\Delta\mu_k^T)\Big) \\ &= \Delta CD_k^{(1,2)}(\frac{1}{2})\end{aligned} \tag{66}$$

$$\begin{aligned}\Delta RD_k^{(1,2)}(\alpha) &= -\frac{1}{2(\alpha-1)}\ln\Big(\frac{|Q_{k,2}|^{\alpha}}{|Q_{k+1,2}|^{\alpha}}\frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|}\frac{|Q_{k,1}|^{1-\alpha}}{|Q_{k+1,1}|^{1-\alpha}}\Big) \\ &+ \frac{\alpha}{2}\Big(\mathrm{Tr}\big(Q_{k+1,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_{k+1}\Delta\mu_{k+1}^T\big) - \mathrm{Tr}\big(Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Delta\mu_k^T\big)\Big)\end{aligned} \tag{67}$$

In addition, given the link between the Rényi divergence and the $\alpha\gamma$-divergence, one gets with γ in the interval $]0,1[$:

$$\begin{aligned}\Delta AG_k^{(1,2)}(\gamma) &= -\frac{1}{2(\gamma-1)\gamma}\ln\Big(\frac{|Q_{k,2}|^{\gamma}}{|Q_{k+1,2}|^{\gamma}}\frac{|Q_{k+1,wei}(1-\gamma,\gamma)|}{|Q_{k,wei}(1-\gamma,\gamma)|}\frac{|Q_{k,1}|^{1-\gamma}}{|Q_{k+1,1}|^{1-\gamma}}\Big) \\ &+ \frac{1}{2}\Big(\mathrm{Tr}\big(Q_{k+1,wei}^{-1}(1-\gamma,\gamma)\Delta\mu_{k+1}\Delta\mu_{k+1}^T\big) - \mathrm{Tr}\big(Q_{k,wei}^{-1}(1-\gamma,\gamma)\Delta\mu_k\Delta\mu_k^T\big)\Big) \\ &= \frac{\Delta RD_k^{(1,2)}(\gamma)}{\gamma}\end{aligned} \tag{68}$$

Note that one has necessarily:

$$\Delta CD_k^{(1,2)}(\alpha) = (1-\alpha)\Delta RD_k^{(1,2)}(\alpha) = \alpha(1-\alpha)\Delta AG_k^{(1,2)}(\alpha) \tag{69}$$

and

$$\Delta BD_k^{(1,2)} = \Delta CD_k^{(1,2)}(\frac{1}{2}) = \frac{1}{2}\Delta RD_k^{(1,2)}(\frac{1}{2}) = \frac{1}{4}\Delta AG_k^{(1,2)}(\frac{1}{2}) \tag{70}$$

Consequently, comparing two Gaussian processes using $\Delta CD_k^{(1,2)}(\alpha)$, $RD_k^{(1,2)}(\alpha)$ or $\Delta AG_k^{(1,2)}(\alpha)$ is equivalent. Moreover, based on (64), one obtains for the γ-divergence the following increment provided $\gamma > 1$:

$$\begin{aligned}\Delta G_k^{(1,2)}(\gamma) &= \frac{1}{2\gamma(\gamma-1)}\ln\Big(\frac{|Q_{k,2}||Q_{k+1,wei}(\gamma-1,1)|^{\gamma}|Q_{k,1}|^{\gamma-1}}{|Q_{k+1,2}||Q_{k,wei}(\gamma-1,1)|^{\gamma}|Q_{k+1,1}|^{\gamma-1}}\Big) \\ &+ \frac{1}{2}\Big(\mathrm{Tr}(Q_{k+1,wei}^{-1}(\gamma-1,1)\Delta\mu_{k+1}\Delta\mu_{k+1}^T) - \mathrm{Tr}(Q_{k,wei}^{-1}(\gamma-1,1)\Delta\mu_k\Delta\mu_k^T)\Big) \\ &- \frac{\ln\gamma}{2(\gamma-1)}\end{aligned} \tag{71}$$

When looking at this increment, we can see that a general expression can encompass all these results. If $\Delta D_k^{(1,2)}(a,b)$ denotes the divergence rate, where a

and b are chosen so that $Q_{k,wei}(b,a)$ and $Q_{k+1,wei}(b,a)$ are definite non negative, one has:

$$\begin{aligned}\Delta D_k^{(1,2)}(a,b) \propto & \frac{1}{2ab(a+b)} \ln\left(\frac{|Q_{k,2}|^a |Q_{k+1,wei}(b,a)|^{a+b} |Q_{k,1}|^b}{|Q_{k+1,2}|^a |Q_{k,wei}(b,a)|^{a+b} |Q_{k+1,1}|^b}\right) \\ & + \frac{1}{2}\left(\mathrm{Tr}(Q_{k+1,wei}^{-1}(b,a)\Delta\mu_{k+1}\Delta\mu_{k+1}^T) - \mathrm{Tr}(Q_{k,wei}^{-1}(b,a)\Delta\mu_k\Delta\mu_k^T)\right) \\ & - \frac{\ln(a+b)}{2ab}\end{aligned} \tag{72}$$

Thus, one has [10]:

$$\begin{cases} \Delta AG_k^{(1,2)}(\gamma) = \Delta D_k^{(1,2)}(1-\gamma,\gamma) \text{ with } \gamma \in]0,1[\\ \Delta RD_k^{(1,2)}(\alpha) \propto \Delta D_k^{(1,2)}(1-\alpha,\alpha) \text{ with } \alpha \in]0,1[\\ \Delta CD_k^{(1,2)}(\alpha) \propto \Delta D_k^{(1,2)}(1-\alpha,\alpha) \text{ with } \alpha \in]0,1[\\ \Delta BD_k^{(1,2)} \propto \Delta D_k^{(1,2)}(\frac{1}{2},\frac{1}{2}) \\ \Delta G_k^{(1,2)}(\gamma) = \Delta D_k^{(1,2)}(1,\gamma-1) \text{ with } \gamma > 1 \end{cases} \tag{73}$$

4.2. Behavior of the other divergences as functions of k

4.2.1. β-divergence

Concerning the analytical expression of the β-divergence increment, we propose to write it this way:

$$\Delta BeD_k^{(1,2)}(\beta) = \Delta BeD_{k,1}^{(1,2)}(\beta) + \Delta BeD_{k,2}^{(1,2)}(\beta) + \Delta BeD_{k,3}^{(1,2)}(\beta) \tag{74}$$

with:

$$\Delta BeD_{k,1}^{(1,2)}(\beta) = \frac{1}{(\beta-1)\beta^{\frac{k}{2}+1}(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}\left(\frac{1}{\beta^{\frac{1}{2}}(2\pi)^{\frac{(\beta-1)}{2}}}\frac{|Q_{k,1}|^{\frac{(\beta-1)}{2}}}{|Q_{k+1,1}|^{\frac{(\beta-1)}{2}}} - 1\right) \tag{75}$$

$$\Delta BeD_{k,2}^{(1,2)}(\beta) = \frac{1}{\beta^{\frac{k}{2}+1}(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,2}|^{\frac{(\beta-1)}{2}}}\left(\frac{1}{\beta^{\frac{1}{2}+1}(2\pi)^{\frac{(\beta-1)}{2}}}\frac{|Q_{k,2}|^{\frac{(\beta-1)}{2}}}{|Q_{k+1,2}|^{\frac{(\beta-1)}{2}}} - 1\right) \tag{76}$$

[10] For information, the increment of the KL divergence satisfies:

$$\begin{aligned}\Delta KL_k^{(1,2)} = \frac{1}{2}\Big[& Tr(Q_{k+1,2}^{-1}(Q_{k+1,1}+\Delta\mu_{k+1}\Delta\mu_{k+1}^T) - Tr(Q_{k,2}^{-1}(Q_{k,1}+\Delta\mu_k\Delta\mu_k^T)) \\ & -1 - \ln\frac{|Q_{k+1,1}||Q_{k,2}|}{|Q_{k,1}||Q_{k+1,2}|}\Big]\end{aligned}$$

$$\Delta BeD_{k,3}^{(1,2)}(\beta) = -\frac{1}{\beta-1}\frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}\times \tag{77}$$
$$\exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)\Big[\frac{|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}{(2\pi)^{\frac{(\beta-1)}{2}}|Q_{k+1,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k+1,2}|^{\frac{\beta-2}{2}}}$$
$$\exp\big(-\frac{(\beta-1)}{2}(\Delta\mu_{k+1}^T Q_{k+1,wei}^{-1}(\beta-1,1)\Delta\mu_{k+1}-\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k)\big)-1\Big]$$

It should be noted that when the processes are both zero-mean or when the difference mean is null, the above expression reduces to:

$$\Delta BeD_{k,3}^{(1,2)}(\beta) = -\frac{1}{\beta-1}\frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}} \tag{78}$$
$$\times\left(\frac{|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}{(2\pi)^{\frac{(\beta-1)}{2}}|Q_{k+1,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k+1,2}|^{\frac{\beta-2}{2}}}-1\right)$$

4.2.2. Other cases

Due to the definition of some divergences like the α-divergence, the Tsallis divergence, the reversed Tsallis divergence, the squared Hellinger distance and the SM divergence, studying the increment when k increases is not necessarily useful as these divergences as functions of k cannot tend to an asymptote. For this reason, let us focus our attention on the normalized increment[11].

Let us start by the normalized increment of the α-divergence by expressing $\Delta A_k^{(1,2)}(\alpha) = A_{k+1}^{(1,2)}(\alpha) - A_k^{(1,2)}(\alpha)$. Using (30) and after simplifications, one has:

$$\Delta A_k^{(1,2)}(\alpha) = -\frac{\exp\left[(\alpha-1)RD_k^{(1,2)}(\alpha)\right]}{\alpha(1-\alpha)}\Big(\exp\left[(\alpha-1)\Delta RD_k^{(1,2)}(\alpha)\right]-1\Big) \tag{79}$$

Therefore, the normalized increment of this divergence is equal to:

$$\frac{\Delta A_k^{(1,2)}(\alpha)}{A_k^{(1,2)}(\alpha)} = \frac{\exp\left[(\alpha-1)\Delta RD_k^{(1,2)}(\alpha)\right]-1}{1-\exp\left[-(\alpha-1)RD_k^{(1,2)}(\alpha)\right]} \tag{80}$$

[11] The normalized increment is not useful for the divergences studied in the previous section. Indeed, one has $\lim_{k\to+\infty} RD_k^{(1,2)} = +\infty$. As studied in a recent paper, $\Delta RD_k^{(1,2)}$ is finite in most of the cases. Therefore, $\lim_{k\to+\infty}\frac{\Delta RD_k^{(1,2)}}{RD_k^{(1,2)}} = 0$. Similar comments can be done for the Chernoff divergence, the Bhattacharyya distance, the γ-divergence and the $\alpha\gamma$-divergence and the KL divergence.

Concerning the normalized increment of the squared Hellinger distance, the Tsallis and the reversed Tsallis divergence, one obtains:

$$\frac{\Delta SH_k^{(1,2)}}{SH_k^{(1,2)}} = \frac{\Delta A_k^{(1,2)}(\frac{1}{2})}{A_k^{(1,2)}(\frac{1}{2})} = \frac{\exp\left[\frac{1}{2}(\Delta RD_k^{(1,2)}(\frac{1}{2})\right] - 1}{1 - \exp\left[\frac{1}{2}RD_k^{(1,2)}(\frac{1}{2})\right]} \tag{81}$$

$$\frac{\Delta T_k^{(1,2)}(\alpha)}{T_k^{(1,2)}(\alpha)} = \frac{\Delta A_k^{(1,2)}(\alpha)}{A_k^{(1,2)}(\alpha)} \tag{82}$$

$$\frac{\Delta rT_k^{(1,2)}(\alpha)}{rT_k^{(1,2)}(\alpha)} = \frac{\Delta A_k^{(1,2)}(1-\alpha)}{A_k^{(1,2)}(1-\alpha)} \tag{83}$$

One can see that analyzing the normalized increment of the Tsallis divergence and the α-divergence leads to the same result. Concerning the reversed Tsallis divergence of order α, this amounts to looking at the behaviour of the normalized increment of the α-divergence of order $1-\alpha$. When $\alpha = \frac{1}{2}$, all these normalized increments are equal and correspond to the normalized increment of the squared Hellinger distance.

Finally, let us define the normalized increment of the SM divergence that is equal to:

$$\frac{\Delta SM_k^{(1,2)}(\alpha,\beta)}{SM_k^{(1,2)}(\alpha,\beta)} = \frac{\exp\left[(\beta-1)\Delta RD_k^{(1,2)}(\alpha)\right] - 1}{1 - \exp\left[-(\beta-1)RD_k^{(1,2)}(\alpha)\right]} \tag{84}$$

Given the above expressions, one can provide the following general expression that encompass the expression of the normalized increment:

$$\frac{\exp\left(c\Delta D_k^{(1,2)}(a,b)\right) - 1}{1 - \exp\left(-cD_k^{(1,2)}(a,b)\right)} \tag{85}$$

Let us address in the next subsection the limit when k tends to infinity.

4.3. Asymptotic behaviour when dealing with NSCE processes

4.3.1. Divergence rate

Let us first look at the divergence rate of the Chernoff, Battacharrya and Rényi divergence. As the processes are zero-mean, $\Delta\mu_k$ and $\Delta\mu_{k+1}$ are null vectors. In addition, using (19), one gets:

$$\Delta CD_k^{(1,2)}(\alpha) = \frac{1}{2}\ln\left(\frac{(1-\alpha)\sigma_1^2 + \alpha\sigma_2^2}{\sigma_1^{2(1-\alpha)}\sigma_2^{2\alpha}}\right) = \frac{1}{2}\ln\left(\frac{(1-\alpha) + \alpha\frac{\sigma_2^2}{\sigma_1^2}}{\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\alpha}}\right) \tag{86}$$

and

$$\Delta BD_k^{(1,2)} = \frac{1}{2}\ln\Big(\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1\sigma_2}\Big) = \frac{1}{2}\ln\Big(\frac{1}{2}\frac{1+\frac{\sigma_2^2}{\sigma_1^2}}{\big(\frac{\sigma_2^2}{\sigma_1^2}\big)^{\frac{1}{2}}}\Big) = \Delta CD_k^{(1,2)}(\frac{1}{2}) \tag{87}$$

and

$$\begin{aligned}\Delta RD_k^{(1,2)}(\alpha) &= -\frac{1}{2(\alpha-1)}\ln\Big(\frac{(1-\alpha)\sigma_1^2+\alpha\sigma_2^2}{\sigma_1^{2(1-\alpha)}\sigma_2^{2\alpha}}\Big) \\ &= -\frac{1}{2(\alpha-1)}\ln\Big(\frac{(1-\alpha)+\alpha\frac{\sigma_2^2}{\sigma_1^2}}{\big(\frac{\sigma_2^2}{\sigma_1^2}\big)^{\alpha}}\Big)\end{aligned} \tag{88}$$

Moreover, one has for γ in the interval $]0,1[$:

$$\begin{aligned}\Delta AG_k^{(1,2)}(\gamma) &= -\frac{1}{2(\gamma-1)\gamma}\ln\Big(\frac{(1-\gamma)\sigma_1^2+\gamma\sigma_2^2}{\sigma_1^{2(1-\gamma)}\sigma_2^{2\gamma}}\Big) \\ &= -\frac{1}{2(\gamma-1)\gamma}\ln\Big(\frac{(1-\gamma)+\gamma\frac{\sigma_2^2}{\sigma_1^2}}{\big(\frac{\sigma_2^2}{\sigma_1^2}\big)^{\gamma}}\Big)\end{aligned} \tag{89}$$

The divergence rate does not depend on the normalized angular frequencies. Moreover, based on (64), one obtains for the γ-divergence the following increment for $\gamma > 1$:

$$\begin{aligned}\Delta G_k^{(1,2)}(\gamma) &= \frac{1}{2\gamma(\gamma-1)}\ln\Big(\frac{((\gamma-1)\sigma_1^2+\sigma_2^2)^{\gamma}}{\sigma_1^{2(\gamma-1)}\sigma_2^2}\Big) - \frac{\ln\gamma}{2(\gamma-1)} \\ &= \frac{1}{2\gamma(\gamma-1)}\ln\Big(\frac{(\gamma-1+\frac{\sigma_2^2}{\sigma_1^2})^{\gamma}}{\frac{\sigma_2^2}{\sigma_1^2}}\Big) - \frac{\ln\gamma}{2(\gamma-1)}\end{aligned} \tag{90}$$

It should be noted that if the inverse of γ was α, one would get:

$$\begin{aligned}\Delta G_k^{(1,2)}(\frac{1}{\alpha}) &= \frac{1}{2(\frac{1}{\alpha}-1)}\ln\Big(\frac{(\frac{1}{\alpha}-1+\frac{\sigma_2^2}{\sigma_1^2})}{(\frac{\sigma_2^2}{\sigma_1^2})^{\alpha}}\Big) - \frac{\ln\frac{1}{\alpha}}{2(\frac{1}{\alpha}-1)} \\ &= \frac{\alpha}{2(1-\alpha)}\ln\Big(\frac{(1-\alpha)+\alpha\frac{\sigma_2^2}{\sigma_1^2}}{(\frac{\sigma_2^2}{\sigma_1^2})^{\alpha}}\Big) = \frac{\alpha}{1-\alpha}\Delta CD_k^{(1,2)}(\alpha)\end{aligned} \tag{91}$$

Depending on the values of γ and α, the divergence rates of the Chernoff divergence and of the γ-divergence can be related to each other.

$\Delta CD_k^{(1,2)}(\alpha)$, $RD_k^{(1,2)}(\alpha)$, $\Delta AG_k^{(1,2)}(\alpha)$ or $\Delta G_k^{(1,2)}(\gamma)$ only depend on the ratio of the additive-noise variances. Comparing two NSCE using these asymptotic increments is hence not necessarily well-suited as it does not take into account the normalized angular frequency. In a previous paper [18], we suggested using the Jeffreys divergence, the symmetric version of the KL divergence defined as $JD_k^{(1,2)} = \frac{1}{2}\left(KL_k^{(1,2)} + KL_k^{(2,1)}\right)$. We obtained[12]:

$$\Delta JD^{(1,2)} = \frac{1}{2}\left(-1 + \frac{1}{2}\left[\Delta T^{(2,1)} + \Delta T^{(1,2)}\right]\right) \tag{92}$$

with:

$$\Delta T^{(i,i')} = \sum_{m=1}^{M_i} \frac{\gamma_{m,i}}{{\sigma'_i}^2}\left(1 - \sum_{n=1}^{M_{i'}} \delta_{\theta_{m,i},\theta_{n,i'}}\right) + \frac{{\sigma_i}^2}{{\sigma'_i}^2} \tag{93}$$

where $\delta_{\theta_{m,i},\theta_{n,i'}}$ denotes the Kronecker delta. It is equal to 1 if $\theta_{m,i} = \theta_{n,i'}$ and to 0 otherwise. In Appendix C, we have derived the expression of the KL divergence rate when dealing with two NSCE processes. One obtains:

$$\Delta KL^{(1,2)} = \frac{1}{2}\Big[\sum_{m=1}^{M_1} \frac{\gamma_{m,1}}{\sigma_2^2}\left(1 - \sum_{n=1}^{M_2} \delta_{\theta_{m,1},\theta_{n,2}}\right) + \frac{\sigma_1^2}{\sigma_2^2} - 1 - \ln\frac{\sigma_1^2}{\sigma_2^2}\Big] \tag{94}$$

Remark. *At first sight, we could wonder why the KL divergence rate depends on the normalized angular frequency unlike the other divergences. To retrieve this result, starting from the expression of the increment $\Delta RD_k^{(1,2)}(\alpha)$, one cannot directly compute the limit of $\Delta RD_k^{(1,2)}(\alpha)$ when α tends to 1 because the logarithm tends to 0 as well as the denominator $-2(\alpha-1)$ of the expression. Once again, L'Hôspital rule must be used. This means that one has to analyze the limit of the ratio of the derivates of both the logarithm and the denominator $-2(\alpha-1)$ when α tends to 1. The details of the mathematical development are given in Appendix C.2.*

[12] Note the expression that was published had some error and is now corrected in the paper. We apologize for the inconvenience.

4.3.2. Asymptotic behaviour of the β-divergence increment

Given (74) one gets:

$$\Delta BeD^{(1,2)}(\beta) = \Delta BeD_1^{(1,2)}(\beta) + \Delta BeD_2^{(1,2)}(\beta) + \Delta BeD_3^{(1,2)}(\beta) \tag{95}$$

where:

1. The first term is given by:

$$\Delta BeD_1^{(1,2)}(\beta) = \frac{1}{\beta(\beta-1)} \left(\frac{1}{\beta^{\frac{1}{2}}(2\pi)^{\frac{(\beta-1)}{2}}\sigma_1^{\beta-1}} - 1 \right) \times \lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}} \tag{96}$$

   As $|Q_{k,1}| > \sigma_1^{2k}$, $\frac{1}{|Q_{k,1}|^{\frac{(\beta-1)}{2}}} < \frac{1}{\sigma_1^{k(\beta-1)}}$. So, we propose to analyze the following upper bound:

$$\Delta BeD_1^{(1,2)}(\beta) < \frac{1}{\beta(\beta-1)} \left(\frac{1}{\beta^{\frac{1}{2}}(2\pi)^{\frac{(\beta-1)}{2}}\sigma_1^{\beta-1}} - 1 \right) \times \lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}\sigma_1^{k(\beta-1)}} \tag{97}$$

   Then the purpose is to analyze the value that $\beta(2\pi)^{\beta-1}\sigma_1^{2(\beta-1)}$ takes with respect to σ_1 and $\beta > 1$. If $\beta(2\pi)^{\beta-1}\sigma_1^{2(\beta-1)} > 1$ *i.e.* $\sigma_1^2 > \frac{1}{2\pi\beta^{\frac{1}{\beta-1}}}$, the limit $\Delta BeD_1^{(1,2)}(\beta)$ is null. When $\beta = 2$, this amounts to saying that $\Delta BeD_1^{(1,2)}(2)$ is null when $\sigma_1^2 > \frac{1}{4\pi}$. As the analysis is made on the upper bound, this leads to a more restrictive condition that guarantees that $\Delta BeD_1^{(1,2)}(\beta) = 0$. In other words, there may be values of σ_1^2 smaller than $\frac{1}{4\pi}$ that could lead to the same result but we did not find any proof for the moment. An alternative would be to express properly the determinant of $Q_{k,1}$. When k tends to infinity, we can show that $Q_{k,1}$ admits M_1 predominant eigenvalues equal to $\{\gamma_{m,1}k + \sigma_1^2\}_{m=1,\ldots,M_1}$ and $k - M_1$ eigenvalues equal to σ_1^2. Therefore, when k tends to infinity, $|Q_{k,1}| = \sigma_1^{2(k-M_1)} \prod_{m=1}^{M_1} \gamma_{m,1}k + \sigma_1^2$. Therefore (96) becomes:

$$\Delta BeD_1^{(1,2)}(\beta) = \frac{\sigma_1^{2M_1}}{\beta(\beta-1)} \left(\frac{1}{\beta^{\frac{1}{2}}(2\pi)^{\frac{(\beta-1)}{2}}\sigma_1^{\beta-1}} - 1 \right) \times \lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}\sigma_1^{k(\beta-1)}(\prod_{m=1}^{M_1}\gamma_{m,1}k + \sigma_1^2)^{\frac{(\beta-1)}{2}}} \tag{98}$$

In this case, let us search for an upper bound of the denominator. One has:

$$
\begin{aligned}
(\prod_{m=1}^{M_1} \gamma_{m,1}k + \sigma_1^2)^{(\beta-1)} &= \exp\Big((\beta-1)\sum_{m=1}^{M_1} \ln(\gamma_{m,1}k + \sigma_1^2)\Big) \\
&< \exp\Big((\beta-1)\sum_{m=1}^{M_1} (\gamma_{m,1}k + \sigma_1^2)\Big) \\
&< \Big(\exp\big(\beta-1\big)\sum_{m=1}^{M_1} \gamma_{m,1}\Big)\Big)^k \exp\big((\beta-1)M_1\sigma_1^2\big)
\end{aligned}
\tag{99}
$$

One can deduce that:

$$
\begin{aligned}
&\lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}\sigma_1^{k(\beta-1)}(\prod_{m=1}^{M_1}\gamma_{m,1}k+\sigma_1^2)^{\frac{(\beta-1)}{2}}} \\
&= \frac{1}{\exp\big(\frac{(\beta-1)}{2}M_1\sigma_1^2\big)} \lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}\sigma_1^{k(\beta-1)}\Big(\exp\big((\beta-1)\sum_{m=1}^{M_1}\gamma_{m,1}\big)\Big)^{\frac{k}{2}}}
\end{aligned}
\tag{100}
$$

The purpose is then to look at the value taken by $\beta\Big((2\pi)\sigma_1^2\exp\big(\sum_{m=1}^{M_1}\gamma_{m,1}\big)\Big)^{(\beta-1)}$ and see when it is larger than 1. In this case, the limit tends to 0. This constraint takes into account all the parameters of the first process.

2. The second term satisfies:

$$
\begin{aligned}
\Delta BeD_{k,2}^{(1,2)}(\beta) = \frac{1}{\beta}\left(\frac{1}{\beta^{\frac{1}{2}+1}(2\pi)^{\frac{(\beta-1)}{2}}\sigma_2^{(\beta-1)}} - 1\right) \times \\
\lim_{k\to+\infty} \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{(\beta-1)}{2}}|Q_{k,2}|^{\frac{(\beta-1)}{2}}}
\end{aligned}
\tag{101}
$$

Following a similar reasoning than above, if $\beta(2\pi)^{\beta-1}\sigma_2^{2(\beta-1)} > 1$ *i.e.* $\sigma_2^2 > \frac{1}{2\pi\beta^{\frac{1}{\beta-1}}}$, a limit equal to 0 for $\Delta BeD_2^{(1,2)}(\beta)$ is guaranteed. Therefore, $\Delta BeD_2^{(1,2)}(2)$ is equal to 0 when $\sigma_2^2 > \frac{1}{4\pi}$.

3. The last term is given by:

$$
\begin{aligned}
\Delta BeD_{k,3}^{(1,2)}(\beta) = -\frac{1}{\beta-1}\left(\frac{1}{(2\pi)^{\frac{(\beta-1)}{2}}((\beta-1)\sigma_1^2+\sigma_2^2)^{\frac{1}{2}}\sigma_2^{\beta-2}} - 1\right) \\
\times \lim_{k\to+\infty} \frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}
\end{aligned}
\tag{102}
$$

When $(2\pi)^{(\beta-1)}\big((\beta-1)\sigma_1^2+\sigma_2^2\big)\sigma_2^{2(\beta-2)} > 1$, we are sure that $\Delta BeD_3^{(1,2)}(\beta) = 0$. For instance when $\beta = 2$, When $\sigma_1^2+\sigma_2^2 > \frac{1}{2\pi}$, $\Delta BeD_3^{(1,2)}(2) = 0$.

Consequently, the normalized increment $\Delta BeD^{(1,2)}(\beta)$ tends to 0 depending on the values of the variances of the additive noises. Indeed, if both variances satisfy $\sigma_i^2 > \frac{1}{2\pi\beta^{\frac{1}{\beta-1}}}$ for $i = 1, 2$, then the last condition $(2\pi)^{(\beta-1)}\big((\beta-1)\sigma_1^2 + \sigma_2^2\big)\sigma_2^{2(\beta-2)} > 1$ also holds and $\Delta BeD^{(1,2)}(\beta)$ is guaranteed to tend to 0. This means that the convergence tends to a limit. The latter is equal to 0.

As a conclusion of this theoretical sub-section, the KL divergence rate and the Jeffreys divergence rate depend on all the parameters characterizing the NSCE processes. The others either depend on the variances of the additive noises only -like the Chernoff divergence- or tend to 0 in some circumstances like the β-divergence. This confirms the relevance of the KL and the Jeffreys divergence for comparing the NSCE processes. Analyzing the results of different divergences could be also relevant. In the next subsection, let us look at the asymptotic normalized increment of some divergences for which the asymptotic increment tends to infinity.

Divergences	Divergence rates
Chernoff	$\Delta CD_k^{(1,2)}(\alpha) = \frac{1}{2}\ln\left(\frac{(1-\alpha)+\alpha\frac{\sigma_2^2}{\sigma_1^2}}{\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\alpha}}\right)$
Battacharyya	$\Delta BD_k^{(1,2)} = \frac{1}{2}\ln\left(\frac{1}{2}\frac{1+\frac{\sigma_2^2}{\sigma_1^2}}{\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\frac{1}{2}}}\right)$
Rényi	$\Delta RD_k^{(1,2)}(\alpha) = -\frac{1}{2(\alpha-1)}\ln\left(\frac{(1-\alpha)+\alpha\frac{\sigma_2^2}{\sigma_1^2}}{\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\alpha}}\right)$
$\alpha\gamma$-divergence	$\Delta AG_k^{(1,2)}(\gamma) = -\frac{1}{2(\gamma-1)\gamma}\ln\left(\frac{(1-\gamma)+\gamma\frac{\sigma_2^2}{\sigma_1^2}}{\left(\frac{\sigma_2^2}{\sigma_1^2}\right)^{\gamma}}\right)$
γ-divergence	$\Delta G_k^{(1,2)}(\gamma) = \frac{1}{2\gamma(\gamma-1)}\ln\left(\frac{(\gamma-1+\frac{\sigma_2^2}{\sigma_1^2})^{\gamma}}{\frac{\sigma_2^2}{\sigma_1^2}}\right) - \frac{\ln\gamma}{2(\gamma-1)}$
Kullback-Leibler	$\Delta KL^{(1,2)} = \frac{1}{2}\Big[\sum_{m=1}^{M_1}\frac{\gamma_{m,1}}{\sigma_2^2}(1-\sum_{n=1}^{M_2}\delta_{\theta_{m,1},\theta_{n,2}}) + \frac{\sigma_1^2}{\sigma_2^2} - 1 - \ln\frac{\sigma_1^2}{\sigma_2^2}\Big]$
Jeffreys	$\Delta JD^{(1,2)} = \frac{1}{2}\left(-1 + \frac{1}{2}\left[\Delta T^{(2,1)} + \Delta T^{(1,2)}\right]\right)$ with: $\Delta T^{(i,i')} = \sum_{m=1}^{M_i}\frac{\gamma_{m,i}}{\sigma_i'^2}(1-\sum_{n=1}^{M_{i'}}\delta_{\theta_{m,i},\theta_{n,i'}}) + \frac{\sigma_i}{\sigma_i'^2}$

Table 1: Divergences and corresponding divergence rates

4.3.3. Asymptotic normalized increment

Concerning the asymptotic increment of the α-divergence, as the Rényi divergence tends to infinity when the processes are dissimilar and k tends to infinity, the α-divergence tends to $\frac{1}{\alpha(1-\alpha)}$ due to (30). Therefore, the asymptotic increment $\Delta A^{(1,2)}$ tends to 0. Similarly, the squared Hellinger distance tends to 1 when k increases due to (33). Therefore its asymptotic increment tends to 0. Moreover, the Tsallis and reversed Tsallis divergences tend to $\frac{1}{1-\alpha}$ when k increases due to (30) and (34). Therefore, the asymptotic increments $\Delta T^{(1,2)}$ and $\Delta rT^{(1,2)}$ also tend to 0.

$$\begin{cases} \Delta A^{(1,2)}(\alpha) = 0 \text{ and } \lim\limits_{k\to+\infty} \frac{\Delta A_k^{(1,2)}(\alpha)}{A_k^{(1,2)}(\alpha)} = 0 \text{ for } \alpha \in]0,1[\\ \Delta SH^{(1,2)} = 0 \text{ and } \lim\limits_{k\to+\infty} \frac{\Delta SH_k^{(1,2)}}{SH_k^{(1,2)}} = 0 \\ \Delta T^{(1,2)}(\alpha) = 0 \text{ and } \lim\limits_{k\to+\infty} \frac{\Delta T_k^{(1,2)}(\alpha)}{T_k^{(1,2)}(\alpha)} = 0 \text{ for } \alpha \in]0,1[\\ \Delta rT^{(1,2)}(\alpha) = 0 \text{ and } \lim\limits_{k\to+\infty} \frac{\Delta rT_k^{(1,2)}(\alpha)}{rT_k^{(1,2)}(\alpha)} = 0 \text{ for } \alpha \in]0,1[\end{cases} \tag{103}$$

As a consequence, whatever the processes to be compared, these four asymptotic increments and consequently the asymptotic normalized increments always provide the same values that are all null. Therefore, using these features to compare NSCE processes does not bring in much.

As mentioned in section 3.1, $\lim\limits_{\beta\to 1} \Delta SM_k^{(1,2)} = \Delta RD_k^{(1,2)}$ and $\lim\limits_{\beta\to 1} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = \frac{\Delta RD_k^{(1,2)}}{RN_k^{(1,2)}}$. In addition, one has:

$$\lim_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = \frac{\exp\left[(\beta-1)\Delta RD^{(1,2)}\right] - 1}{1 - \lim\limits_{k\to+\infty} \exp\left[-(\beta-1)RD_k^{(1,2)}\right]} \tag{104}$$

As $RD_k^{(1,2)} \geq 0$ necessarily increases when k increases, $\lim\limits_{k\to+\infty} \exp\left[(1-\beta)RD_k^{(1,2)}\right]$ depends on the values of β. This leads to three cases:

$$\begin{cases} \lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = 0 \text{ if } 0 < \beta < 1 \\ \lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = \lim\limits_{k\to+\infty} \frac{\Delta RD_k^{(1,2)}}{RD_k^{(1,2)}} = 0 \text{ when } \beta \text{ tends to } 1 \\ \lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = \exp\left[(\beta-1)\Delta RD^{(1,2)}\right] - 1 \text{ if } \beta > 1 \end{cases} \tag{105}$$

Selecting β in the interval $]0,1[$ does not help the practitioner to compare the NSCE processes if he/she uses the asymptotic normalized increment because it

is always equal to 0. Therefore, β must be chosen larger than 1. In this case and for a given value of α, as $\exp\left[(\beta-1)\Delta RD^{(1,2)}\right]$is an increasing function of β when $\beta > 1$, $\lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}}$ is also an increasing function of β . However, when the ratio of the additive-noise variances is equal to 1, the divergence rate of the Rényi divergence is equal to 0. Consequently, $\exp\left[(\beta-1)\Delta RD^{(1,2)}\right] = 1$ and $\lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = 0$.

5. Illustrations

In this section and given the length of the paper, we present the way the divergences evolve for a single example. The reader can easily refer to the expressions of the increments given in the previous section and simulate different cases.

The two processes are characterized by the following parameters: the first NSCE process is a sum of three complex exponentials whose normalized angular frequencies are equal to $\theta_{1,1} = \frac{2\pi}{5}$, $\theta_{2,1} = \frac{\pi}{2}$ and $\theta_{3,1} = \frac{\pi}{3}$. The variances of the magnitudes are respectively equal to $\gamma_{1,1} = 5$, $\gamma_{2,1} = 10$ and $\gamma_{3,1} = 15$. The variance of the additive noise is equal to $\sigma_1^2 = 2$. The second NSCE process is a sum of two complex exponentials whose first normalized angular frequency is equal to $\theta_{1,2} = \frac{\pi}{2}$. Regarding $\theta_{2,2}$, it is first set at $\frac{\pi}{15}$. Then, we will analyze the case where $\theta_{2,2}$ is modified and becomes equal to $\frac{\pi}{10}$ and $\frac{\pi}{5}$. The variances of the magnitudes are respectively equal to 30 and 20. The variance of the additive noise is equal to $\sigma_2^2 = 1$. Concerning the divergences, α is set at 0.7. β is equal to 3. γ is set at 2.

If the divergence rate is not null, this means that, when k increases and tends to infinity, the divergence can be approximated by an affine function whose slope is the divergence rate. Depending on the value of the slope, the shape of the divergence as a function k may be different. Thus, when looking at Figure 2, the increment Kullback-Leibler divergence converges to the KL divergence rate when k increases. The variations around the divergence rate becomes smaller and smaller when k increases, but the order of magnitude is always the same.

In the illustration, the increment is rapidly close to the divergence rate, giving the feeling to the practitioner the KL divergence -as a function of k- can be approximated by an affine function even when k takes small values. Moreover, when $\theta_{2,2}$ is modified, this has an impact on the variations of the increment around the divergence rate and the speed of convergence toward the divergence rate. However, as the orders of magnitudes of the increments for the different values of $\theta_{2,2}$ are more or less the same, the values of the divergence are quite similar.

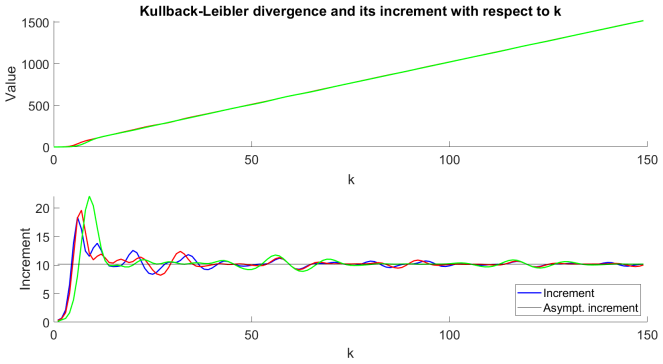

Figure 2: Kullback-Leibler divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

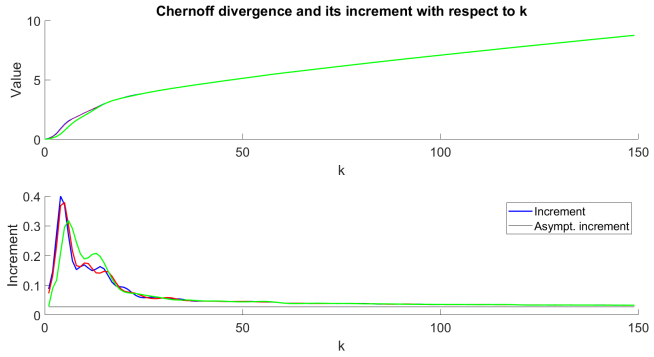

Figure 3: Chernoff divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

When the divergence rate is small and close to zero, the increment of the diver-

gence rate which was not necessarily small when k was small, is close to zero when k becomes larger. The value of the increment varies much. Given the numerical example we chose, this explains the shapes of the Chernoff divergence in Figure 3 where the maximal value of the increment is almost 0.4 but the rate is equal to 0.0272 (*i.e.* 15 times less). Therefore, the Chernoff divergence, which can be approximated as an affine function whose slope is 0.0272 when k tends to infinity, slowly increases and is not bounded.

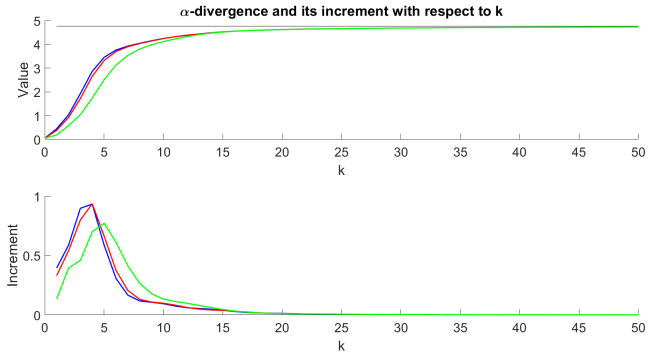

Figure 4: α-divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

Regarding the α-divergence in Figure 4, it is confirmed that the α-divergence tends to $\frac{1}{\alpha(1-\alpha)} \approx 4.76$ while the asymptotic increment $\Delta A^{(1,2)}$ tends to 0.

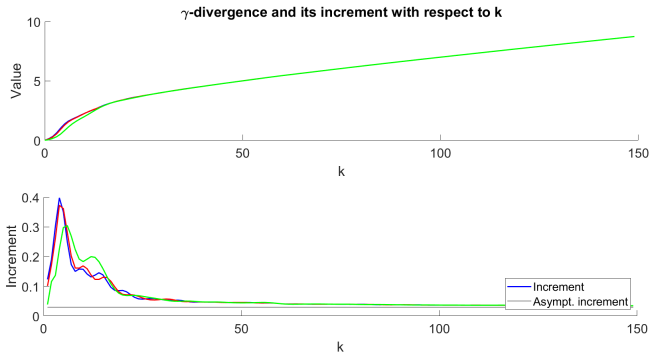

Figure 5: γ-divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

In Figure 5, the γ-divergence slowly increases since its divergence rate is equal to 0.0294. It is not bounded.

Let us now look at the $\alpha\gamma$-divergence presented in Figure 6 below. Since $\Delta AG_k^{(1,2)}(\gamma) = \frac{\Delta CD_k^{(1,2)}(\gamma)}{\gamma(1-\gamma)}$, it is no coincidence that the curves in Figure 6 look like those in Figure 3, except for a multiplicative factor.

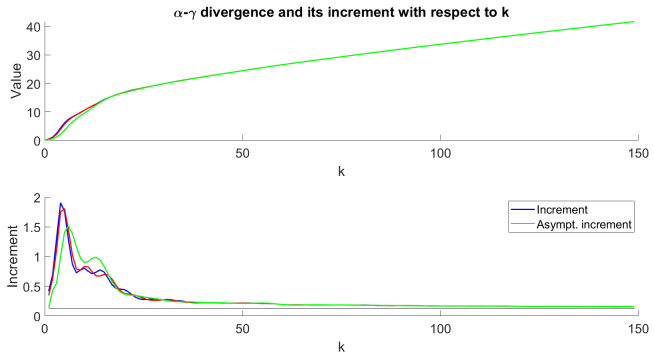

Figure 6: $\alpha\gamma$-divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

In Figures 7 and 8, as expected, the Tsallis and reversed Tsallis divergences tend to $\frac{1}{1-\alpha} \approx 3.33$ when k increases. Therefore, the asymptotic increments $\Delta T^{(1,2)}$ and $\Delta rT^{(1,2)}$ tend to 0.

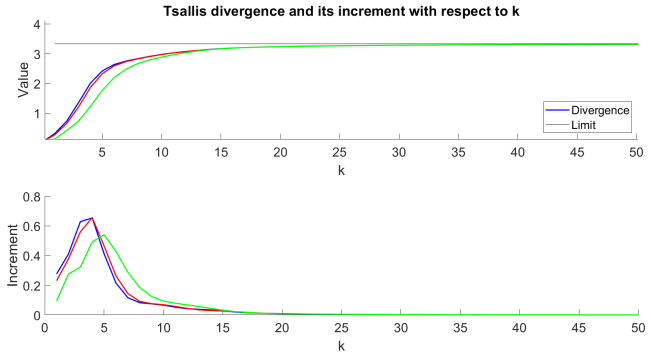

Figure 7: Tsallis divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

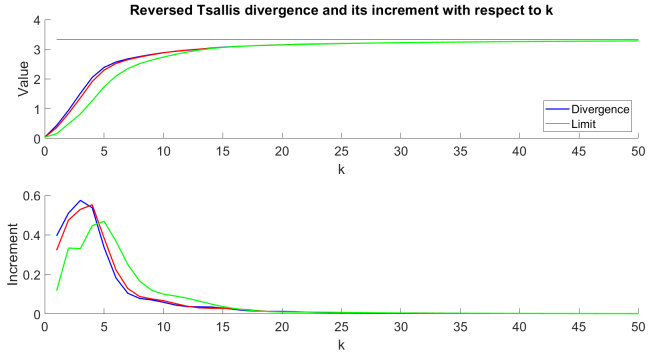

Figure 8: Reversed Tsallis divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

In Figure 9, one can see that the squared Hellinger distance tends to 1 when k increases, as mentioned in the theoretical part. Therefore its asymptotic increment tends to 0, as expected.

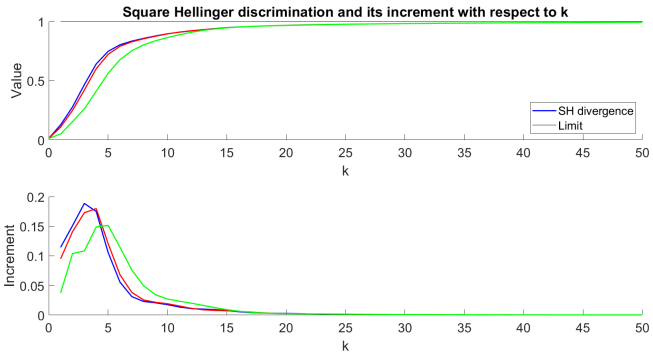

Figure 9: Squared Hellinger distance evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

Let us end up this section with the β-divergence in Figure 10. Its shape is different from what we observed with the other divergences. The increment is negative and the divergence tends to decrease.

Regarding the Sharma-Mittal divergence given in Figure 11, the normalized increment converges to the asymptotic normalized increment equal to 0.1989. This value coincides to the theoretical result we obtained when $\beta > 1$ in

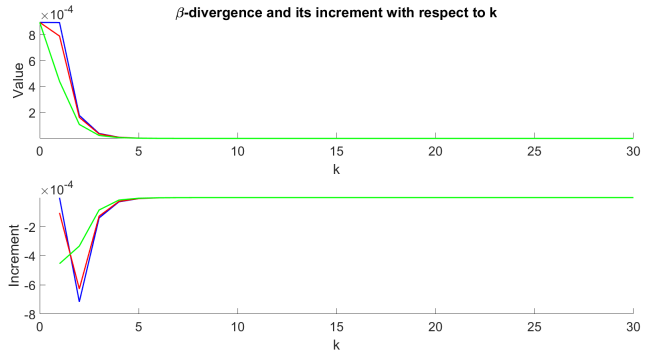

Figure 10: β-divergence evolution and increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

(105). Indeed as $\Delta CD^{(1,2)} = 0.0272$, $\Delta RD^{(1,2)} = -\frac{1}{\alpha-1}\Delta CD^{(1,2)} = 0.0907$ and $\lim\limits_{k\to+\infty} \frac{\Delta SM_k^{(1,2)}}{SM_k^{(1,2)}} = \exp[(\beta-1)\Delta RD^{(1,2)}] - 1 = 0.1989$.

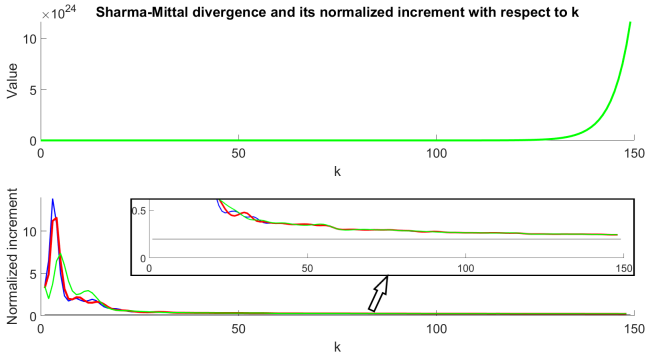

Figure 11: Sharma-Mittal divergence evolution and normalized increment evolution for the given example. Blue curve for $\theta_{2,2} = \frac{\pi}{15}$, red curve for $\theta_{2,2} = \frac{\pi}{10}$ and green curve for $\theta_{2,2} = \frac{\pi}{5}$.

6. Conclusions and perspectives

In this paper, we have deduced expressions of various divergences between Gaussian processes modeled by sums of complex exponentials disturbed by additive noises. Among all the divergences that have been studied, it appears the KL and Jeffreys divergence rate have the advantage of depending on the model parameters. Many others only depend on the ratio of the additive-noise variances. Sharma-Mittal divergence can be also of interest, but the asymptotic normalized increment must be considered in this case except when the variances of the additive noises are the same. The illustrations confirm the theoretical analysis. As a short-term perspective, we plan to analyze other divergences and other types of random processes.

Appendix A. Derivation of quantities useful to get the analytical expression of the β-divergence and the γ-divergence in the Gaussian case

Appendix A.1. Expression of $\int_{X_k} p_1^{\beta}(X_k)dX_k$

Given (49), one has:

$$
\begin{aligned}
I_2(p_1,\beta) &= \int_{X_k} p_1^{\beta}(X_k)dX_k \qquad (A.1)\\
&= \int_{X_k} \frac{1}{(2\pi)^{\frac{k\beta}{2}}|Q_{k,1}|^{\frac{\beta}{2}}} \exp\big(-\frac{\beta}{2}[X_k-\mu_{k,1}]^T Q_{k,1}^{-1}[X_k-\mu_{k,1}]\big)dX_k\\
&= \frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}\times\\
&\int_{X_k} \frac{1}{(2\pi)^{\frac{k}{2}}|Q_{k,1}|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}[X_k-\mu_{k,1}]^T \left(\frac{1}{\beta}Q_{k,1}\right)^{-1}[X_k-\mu_{k,1}]\big)dX_k
\end{aligned}
$$

As the determinant of $\frac{1}{\beta}Q_{k,1}$ is equal to $\frac{1}{\beta^k}|Q_{k,1}|$, the above equality can be rewritten this way:

$$
\begin{aligned}
I_2(p_1,\beta) &= \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}\times \qquad (A.2)\\
&\int_{X_k} \frac{\beta^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}|Q_{k,1}|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}[X_k-\mu_{k,1}]^T \left(\frac{1}{\beta}Q_{k,1}\right)^{-1}[X_k-\mu_{k,1}]\big)dX_k\\
&= \frac{1}{\beta^{\frac{k}{2}}(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}
\end{aligned}
$$

Appendix A.2. Expression of $\int_{X_k} p_1^a(X_k) p_2^b(X_k) dX_k$

Let us now address the expression of the second integral defined by:

$$I_1(p_1, p_2, a, b) = \int \frac{1}{(\sqrt{2\pi})^{k(a+b)} |Q_{k,1}|^{\frac{a}{2}} |Q_{k,2}|^{\frac{b}{2}}} \exp\big(-\frac{1}{2}A\big) \tag{A.3}$$

where $a > 0$ and $b > 0$. We will see why this assumption is required in the rest of the mathematical development. In addition, A can be expressed after some mathematical developments as follows:

$$\begin{aligned} A &= a[X_k - \mu_{k,1}]^T Q_{k,1}^{-1}[X_k - \mu_{k,1}] + b[X_k - \mu_{k,2}]^T Q_{k,2}^{-1}[X_k - \mu_{k,2}] \\ &= X_k^T \big(aQ_{k,1}^{-1} + bQ_{k,2}^{-1}\big) X_k - 2X_k^T \big(aQ_{k,1}^{-1}\mu_{k,1} + bQ_{k,2}^{-1}\mu_{k,2}\big) \\ &+ a\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + b\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} \end{aligned} \tag{A.4}$$

At this stage, let us introduce the following two quantities:

$$Q_{k,a,b}^{-1} = aQ_{k,1}^{-1} + bQ_{k,2}^{-1} \tag{A.5}$$

It should be noted that a and b are necessarily positive to guarantee that $Q_{k,a,b}$ is definite non negative and exploit this property in the remainder of the mathematical development. In addition, one has:

$$Q_{k,a,b}^{-1}\mu_{k,a,b} = aQ_{k,1}^{-1}\mu_{k,1} + bQ_{k,2}^{-1}\mu_{k,2} \tag{A.6}$$

Given (A.5) and (A.6), (A.4) can be rewritten this way:

$$\begin{aligned} A &= (X_k - \mu_{k,a,b})^T Q_{k,a,b}^{-1}(X_k - \mu_{k,a,b}) \\ &- \mu_{k,a,b}^T Q_{k,a,b}^{-1}\mu_{k,a,b} + a\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + b\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} \end{aligned} \tag{A.7}$$

Using (A.7), the next step is to rewrite $I_1(p_1, p_2, a, b)$ by taking into account the fact that $\int \frac{1}{(\sqrt{2\pi})^k |Q_{k,a,b}|^{1/2}} \exp\big(-\frac{1}{2}(X_k - \mu_{k,a,b})^T Q_{k,a,b}^{-1}(X_k - \mu_{k,a,b})\big) \mathrm{d}X_k = 1$. This simplification implicitly implies that $Q_{k,a,b}$ is invertible and has the property of a covariance matrix of a Gaussian process. It is hence assumed to be Hermitian and definite positive. Its inverse $Q_{k,a,b}^{-1}$ has therefore the same properties. When a and b are strictly positive and referring to (A.5), this assumption holds. Indeed, as $x^H Q_{k,1}^{-1} x > 0$ and $x^H Q_{k,2}^{-1} x > 0$ for all non zero column vector x of appropriate size, $x^H Q_{k,a,b}^{-1} x = a x^H Q_{k,1}^{-1} x + b x^H Q_{k,2}^{-1} x > 0$. In this paper,

this constraint on a and b will be enough for the mathematical development we are doing. One obtains:

$$I_1(p_1,p_2,a,b) = \frac{|Q_{k,a,b}|^{1/2}}{(\sqrt{2\pi})^{k(a+b-1)}|Q_{k,1}|^{\frac{a}{2}}|Q_{k,2}|^{\frac{b}{2}}}\times \exp\Big(-\frac{1}{2}(a\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + b\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} - \mu_{k,a,b}^T Q_{k,a,b}^{-1}\mu_{k,a,b})\Big) \tag{A.8}$$

Then, our purpose is to simplify the argument of the complex exponential in (A.8). To this end, let us introduce a new matrix denoted as $Q_{k,wei}(b,a)$ and defined by:

$$Q_{k,wei}(b,a) = bQ_{k,1} + aQ_{k,2} = Q_{k,wei}^T(b,a) \tag{A.9}$$

Then, one has:

$$\begin{aligned} Q_{k,a,b}^{-1} &\underset{(A.5)}{=} aQ_{k,1}^{-1} + bQ_{k,2}^{-1} = Q_{k,1}^{-1}\big(aI_k + bQ_{k,1}Q_{k,2}^{-1}\big) \\ &= Q_{k,1}^{-1}(aQ_{k,2} + bQ_{k,1})Q_{k,2}^{-1} \underset{(A.9)}{=} Q_{k,1}^{-1}Q_{k,wei}(b,a)Q_{k,2}^{-1} \end{aligned} \tag{A.10}$$

Note that:

$$Q_{k,a,b}^{-1} = (Q_{k,a,b}^{-1})^T \text{ and } Q_{k,wei}(b,a) = Q_{k,1}Q_{k,a,b}^{-1}Q_{k,2} = Q_{k,2}Q_{k,a,b}^{-1}Q_{k,1} \tag{A.11}$$

Their determinants hence satisfy:

$$|Q_{k,a,b}| = \frac{|Q_{k,1}||Q_{k,2}|}{|Q_{k,wei}(b,a)|} \tag{A.12}$$

In addition, let us differently express the argument of the complex exponential in (A.8), *i.e.*:

$$B = -\frac{1}{2}\Big(a\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + b\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} - \mu_{k,a,b}^T Q_{k,a,b}^{-1}\mu_{k,a,b})\Big) \tag{A.13}$$

by using (A.6) and by taking into account the fact that $Q_{k,a,b}^{-1}$ is symmetric. This leads to:

$$\begin{aligned} &\mu_{k,a,b}^T Q_{k,a,b}^{-1}\mu_{k,a,b} = \mu_{k,a,b}^T Q_{k,a,b}^{-1}Q_{k,a,b}Q_{k,a,b}^{-1}\mu_{k,a,b} \\ &\underset{(A.6)}{=} \Big(a\mu_{k,1}^T Q_{k,1}^{-1} + b\mu_{k,2}^T Q_{k,2}^{-1}\Big)Q_{k,a,b}\Big(aQ_{k,1}^{-1}\mu_{k,1} + bQ_{k,2}^{-1}\mu_{k,2}\Big) \\ &= a^2\mu_{k,1}^T Q_{k,1}^{-1}Q_{k,a,b}Q_{k,1}^{-1}\mu_{k,1} + b^2\mu_{k,2}^T Q_{k,2}^{-1}Q_{k,a,b}Q_{k,2}^{-1}\mu_{k,2} + 2ab\mu_{k,1}^T Q_{k,1}^{-1}Q_{k,a,b}Q_{k,2}^{-1}\mu_{k,2} \end{aligned} \tag{A.14}$$

Therefore, one has:

$$\begin{aligned} B = -\frac{1}{2}\Big(&a\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} - a^2\mu_{k,1}^T Q_{k,1}^{-1}Q_{k,a,b}Q_{k,1}^{-1}\mu_{k,1} \\ &+ b\mu_{k,2}^T Q_{k,2}^{-1}\mu_{k,2} - b^2\mu_{k,2}^T Q_{k,2}^{-1}Q_{k,a,b}Q_{k,2}^{-1}\mu_{k,2} - 2ab\mu_{k,1}^T Q_{k,1}^{-1}Q_{k,a,b}Q_{k,2}^{-1}\mu_{k,2}\Big) \end{aligned} \tag{A.15}$$

However, given (A.5), $Q^{-1}_{k,wei}(b,a)$ can be expressed using the matrix inversion lemma[13] provided the matrix is invertible. Two expressions can be deduced, depending on the use of the lemma. On the one hand:

$$bQ^{-1}_{k,wei}(b,a) = Q^{-1}_{k,1} - aQ^{-1}_{k,1}Q_{k,a,b}Q^{-1}_{k,1} \tag{A.16}$$

On the other hand:

$$aQ^{-1}_{k,wei}(b,a) = Q^{-1}_{k,2} - bQ^{-1}_{k,2}Q_{k,a,b}Q^{-1}_{k,2} \tag{A.17}$$

By combining (A.11), (A.15), (A.16) and (A.17), one obtains:

$$B = -\frac{ab}{2}\left(\mu^T_{k,1}Q^{-1}_{k,wei}(b,a)\mu_{k,1} + \mu^T_{k,2}Q^{-1}_{k,wei}(b,a)\mu_{k,2}\right) + ab\mu^T_{k,1}Q^{-1}_{k,1}Q_{k,a,b}Q^{-1}_{k,2}\mu_{k,2} \tag{A.18}$$

$$= -\frac{ab}{2}(\mu_{k,1}-\mu_{k,2})^T Q^{-1}_{k,wei}(b,a)(\mu_{k,1}-\mu_{k,2})$$

Finally, by introducing $\Delta\mu_k = \mu_{k,1} - \mu_{k,2}$ and combining (A.8), (A.12) and (A.18), one gets:

$$I_1(p_1,p_2,a,b) = \frac{|Q_{k,1}|^{\frac{1-a}{2}}|Q_{k,2}|^{\frac{1-b}{2}}}{(2\pi)^{\frac{k(a+b-1)}{2}}|Q_{k,wei}(b,a)|^{\frac{1}{2}}}\exp\left(-\frac{ab}{2}\Delta\mu_k^T Q^{-1}_{k,wei}(b,a)\Delta\mu_k\right) \tag{A.19}$$

Therefore, when $a=\alpha>0$ and $b=1-\alpha>0$, meaning $0<\alpha<1$, one has:

$$I_1(p_1,p_2,\alpha,1-\alpha) = \frac{|Q_{k,1}|^{\frac{1-\alpha}{2}}|Q_{k,2}|^{\frac{\alpha}{2}}}{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{1}{2}}}\exp\left(-\frac{\alpha(1-\alpha)}{2}\Delta\mu_k^T Q^{-1}_{k,wei}(1-\alpha,\alpha)\Delta\mu_k\right) \tag{A.20}$$

When $a=1$ and $b=\beta-1>0$, meaning $\beta>1$, this leads to:

$$I_1(p_1,p_2,1,\beta-1) = \frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}} \times \exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q^{-1}_{k,wei}(\beta-1,1)\Delta\mu_k\right) \tag{A.21}$$

Appendix B. Another way to derive the results in the Gaussian case

Let us rewrite (49) differently:

$$p_1(X_k) = \exp\left(-\frac{k}{2}\ln(2) - \frac{k}{2}\ln(\pi) - \frac{1}{2}\ln(|Q_{k,1}|) - \frac{1}{2}[X_k-\mu_{k,1}]^T Q^{-1}_{k,1}[X_k-\mu_{k,1}]\right) \tag{B.1}$$

[13] In the current case, it is given by $(U+V)^{-1} = U^{-1} - U^{-1}(V^{-1}+U^{-1})^{-1}U^{-1}$.

This leads to:

$$p_1(X_k) = \exp\Big(-\frac{k}{2}\ln(\pi) - \frac{k}{2}\ln(2) - \frac{1}{2}\ln(|Q_{k,1}|) - \frac{1}{2}\mu_{k,1}^T Q_{k,1}^{-1}\mu_{k,1} + \mu_{k,1}^T Q_{k,1}^{-1} X_k - \frac{1}{2}X_k^T Q_{k,1}^{-1} X_k\Big) \tag{B.2}$$

or equivalently:

$$p_1(X_k) = \exp\Big(-\frac{k}{2}\ln(\pi) - \frac{1}{2}\ln(2^k|Q_{k,1}|) - \frac{1}{2}Tr(Q_{k,1}^{-1}\mu_{k,1}\mu_{k,1}^T) + (Q_{k,1}^{-1}\mu_{k,1})^T X_k - Tr((2Q_{k,1})^{-1}X_k X_k^T)\Big) \tag{B.3}$$

or

$$p_1(X_k) = \exp\Big(-\frac{k}{2}\ln(\pi) + \frac{1}{2}\ln(|\frac{1}{2}Q_{k,1}^{-1}|) - \frac{1}{2}Tr(Q_{k,1}^{-1}Q_{k,1}Q_{k,1}^{-1}\mu_{k,1}\mu_{k,1}^T) + (Q_{k,1}^{-1}\mu_{k,1})^T X_k - Tr((2Q_{k,1})^{-1}X_k X_k^T)\Big) \tag{B.4}$$

The above expression can be rewritten as:

$$p_1(X_k) = \exp\Big(-\frac{k}{2}\ln(\pi) + \frac{1}{2}\ln(|\frac{1}{2}Q_{k,1}^{-1}|) - \frac{1}{4}Tr(2Q_{k,1}Q_{k,1}^{-1}\mu_{k,1}(Q_{k,1}^{-1}\mu_{k,1})^T) + (Q_{k,1}^{-1}\mu_{k,1})^T X_k - Tr((2Q_{k,1})^{-1}X_k X_k^T)\Big) \tag{B.5}$$

By introducing the scalar product between two symmetric positive definite matrices X and Y as the trace of the product XY *i.e.* $< X, Y >= Tr(XY)$, the so-called sufficient statistics $t(X_k) = \big(X_k, -X_k X_k^T\big)$ and what is called the natural parameters stored in Θ which is the set $(\theta = Q_{k,1}^{-1}\mu_{k,1}, \phi = (2Q_{k,1})^{-1})$, we can conclude that $(Q_{k,1}^{-1}\mu_{k,1})^T X_k - Tr((2Q_{k,1})^{-1}X_k X_k^T)$ can be seen as the scalar product between the sufficient statistics and the natural parameter Θ, *i.e.* $< t(X_k), \Theta >$. In addition, what is known as the log normalizer can be expressed as:

$$F(\Theta) = \frac{1}{4}Tr(\phi^{-1}\theta\theta^T) - \frac{1}{2}\ln(|\phi|) + \frac{k}{2}\ln(\pi) \tag{B.6}$$

Finally by setting the so-called carrier measure $k(X_k) = 0$, one has:

$$p_1(X_k, \Theta) = \exp(< t(X_k), \Theta > -F(\Theta) + k(X_k)) \tag{B.7}$$

In [31] when dealing with pdfs belonging to the exponential family, respectively characterized by the set of natural parameters $\Theta_1 = (\theta = Q_{k,1}^{-1}\mu_{k,1}, \phi = (2Q_{k,1})^{-1})$ and $\Theta_2 = (\theta = Q_{k,2}^{-1}\mu_{k,2}, \phi = (2Q_{k,2})^{-1})$, Nielsen provided a closed

form expression of the following integral when $a+b=1$:

$$I_1(p_1,p_2,a,b)=\int_{X_k}p_1^a(X_k)p_2^b(X_k)dX_k \tag{B.8}$$
$$=\exp\left(F\left(a\theta_1+b\theta_2\right)-aF(\theta_1)-bF(\theta_2)\right)$$

Therefore, the Chernoff coefficient can be deduced as follows:

$$C_\alpha(p_1\left(X_k\right),p_2\left(X_k\right))=\int_{X_k}p_1^\alpha\left(X_k\right)p_2^{1-\alpha}\left(X_k\right)\mathrm{d}X_k \tag{B.9}$$
$$=I_1(p_1,p_2,\alpha,1-\alpha)=\exp\left(F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)-\alpha F(\Theta_1)-(1-\alpha)F(\Theta_2)\right)$$

In this case, the following quantities must be taken into account to calculate $F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)$:

$$\begin{cases}\theta_{equ}(\alpha,1-\alpha) & =\alpha Q_{k,1}^{-1}\mu_{k,1}+(1-\alpha)Q_{k,2}^{-1}\mu_{k,2}\\ \phi_{equ}(\alpha,1-\alpha) & =\frac{1}{2}\left(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1}\right)=\frac{1}{2}Q_{k,1}^{-1}\left(\alpha Q_{k,2}+(1-\alpha)Q_{k,1}\right)Q_{k,2}^{-1}\\ & =\frac{1}{2}Q_{k,1}^{-1}Q_{k,wei}(1-\alpha,\alpha)Q_{k,2}^{-1}\end{cases} \tag{B.10}$$

It should be noted that the inverse of $\phi_{equ}(\alpha,1-\alpha)=\frac{1}{2}\left(\alpha Q_{k,1}^{-1}+(1-\alpha)Q_{k,2}^{-1}\right)$ can be expressed in three different ways. Among them, the inversion matrix lemma can be considered twice. This leads to:

$$\phi_{equ}^{-1}(\alpha,1-\alpha)=2Q_{k,2}Q_{k,wei}^{-1}(1-\alpha,\alpha)Q_{k,1} \tag{B.11}$$
$$=\frac{2}{\alpha}\left(Q_{k,1}-(1-\alpha)Q_{k,1}Q_{k,wei}^{-1}(1-\alpha,\alpha)Q_{k,1}\right)$$
$$=\frac{2}{1-\alpha}\left(Q_{k,2}-\alpha Q_{k,2}Q_{k,wei}^{-1}(1-\alpha,\alpha)Q_{k,2}\right)$$

Therefore, given (B.6), the log normalizer $F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)$ is equal to:

$$F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)=\frac{1}{2}\ln(2^k\frac{|Q_{k,1}||Q_{k,2}|}{|Q_{k,wei}(1-\alpha,\alpha)|})+\frac{k}{2}\ln(\pi) \tag{B.12}$$
$$+\frac{1}{4}\left(\left(\alpha Q_{k,1}^{-1}\mu_{k,1}+(1-\alpha)Q_{k,2}^{-1}\mu_{k,2}\right)^T\phi_{equ}^{-1}\left(\alpha Q_{k,1}^{-1}\mu_{k,1}+(1-\alpha)Q_{k,2}^{-1}\mu_{k,2}\right)\right)$$
$$=\frac{1}{2}\ln(\frac{|Q_{k,1}||Q_{k,2}|}{|Q_{k,wei}(1-\alpha,\alpha)|}+\frac{k}{2}\ln(2)+\frac{k}{2}\ln(\pi)$$
$$+\frac{1}{4}\Big(2\alpha(1-\alpha)\mu_{k,2}^TQ_{k,2}^{-1}\phi_{equ}^{-1}Q_{k,1}^{-1}\mu_{k,1}+\alpha^2\mu_{k,1}^TQ_{k,1}^{-1}\phi_{equ}^{-1}Q_{k,1}^{-1}\mu_{k,1}$$
$$+(1-\alpha)^2\mu_{k,2}^TQ_{k,2}^{-1}\phi_{equ}^{-1}Q_{k,2}^{-1}\mu_{k,2}\Big)$$

At this stage, substituting ϕ_{equ}^{-1} by successively using its three different expres-

sions (B.11), one gets:

$$F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)=\frac{1}{2}\ln(\frac{|Q_{k,1}||Q_{k,2}|}{|Q_{k,wei}(1-\alpha,\alpha)|})+\frac{k}{2}\ln(2\pi) \tag{B.13}$$
$$-\frac{1}{2}\alpha(1-\alpha)\Delta\mu_k Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k+\frac{\alpha}{2}\mu_{k,1}^T Q_{k,1}\mu_{k,i}+\frac{(1-\alpha)}{2}\mu_{k,2}^T Q_{k,1}\mu_{k,2}$$

In addition, for $i=1,2$, one has:

$$F(\Theta_i)=\frac{1}{2}\mu_{k,i}^T Q_{k,i}\mu_{k,i}+\frac{1}{2}\ln(|Q_{k,i}|)+\frac{k}{2}\ln(2\pi) \tag{B.14}$$

Therefore, $\exp(F\left(\alpha\Theta_1+(1-\alpha)\Theta_2\right)-\alpha F(\Theta_1)-(1-\alpha)F(\Theta_2))$ is equal to:

$$C_\alpha(p_1\left(X_k\right),p_2\left(X_k\right))=\int_{X_k}p_1^\alpha\left(X_k\right)p_2^{1-\alpha}\left(X_k\right)\mathrm{d}X_k \tag{B.15}$$
$$=\frac{|Q_{k,1}|^{\frac{1-\alpha}{2}}|Q_{k,2}|^{\frac{\alpha}{2}}}{|Q_{k,wei}(1-\alpha,\alpha)|^{\frac{1}{2}}}\exp\left(-\frac{\alpha(1-\alpha)}{2}\Delta\mu_k Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\right)$$

One hence retrieves the result given in (A.20).

Appendix C. Derivation of the KL divergence for NSCE processes. Proof of the link with the Rényi divergence

Appendix C.1. Derivation of the KL divergence rate

Given the expression of the KL mentioned in the footnote 9 of the paper, the increment of the divergence satisfies:

$$\Delta KL_k^{(1,2)}=\frac{1}{2}\Big[\mathrm{Tr}(Q_{k+1,2}^{-1}Q_{k+1,1})-\mathrm{Tr}(Q_{k,2}^{-1}Q_{k,1})-1-\ln\frac{|Q_{k+1,1}||Q_{k,2}|}{|Q_{k,1}||Q_{k+1,2}|} \tag{C.1}$$
$$+Tr(Q_{2,k+1}^{-1}\Delta\mu_{k+1}\Delta\mu_{k+1}^T)-Tr(Q_{2,k}^{-1}\Delta\mu_k\Delta\mu_k^T)\Big]$$

As the processes under study are zero-mean, the above expression reduces to:

$$\Delta KL_k^{(1,2)}=\frac{1}{2}\Big[\mathrm{Tr}(Q_{k+1,2}^{-1}Q_{k+1,1})-\mathrm{Tr}(Q_{k,2}^{-1}Q_{k,1})-1-\ln\frac{|Q_{k+1,1}||Q_{k,2}|}{|Q_{k,1}||Q_{k+1,2}|}\Big] \tag{C.2}$$

However, the limit of the logarithm of the determinant ratio has been addressed in (19). Thus, one has:

$$\lim_{k\to+\infty}\ln\frac{|Q_{k+1,1}||Q_{k,2}|}{|Q_{k,1}||Q_{k+1,2}|}=\ln\frac{\sigma_1^2}{\sigma_2^2} \tag{C.3}$$

Moreover, the limit of the trace difference was already addressed in one of our previous papers [19]. This led to:

$$\lim_{k\to+\infty}\mathrm{Tr}(Q_{k+1,2}^{-1}Q_{k+1,1})=\Delta T^{(2,1)} \tag{C.4}$$

where:

$$\Delta T^{(i,i')} = \sum_{m=1}^{M_i} \frac{\gamma_{m,i}}{\sigma_i'^2}(1 - \sum_{n=1}^{M_{i'}} \delta_{\theta_{m,i},\theta_{n,i'}}) + \frac{\sigma_i^2}{\sigma_i'^2} \tag{C.5}$$

with $\delta_{\theta_{m,i},\theta_{n,i'}}$ the Kronecker delta. It is equal to 1 if $\theta_{m,i} = \theta_{n,i'}$ and to 0 otherwise. Note that we take advantage of this paper to fix a typo mistake that appeared in [19]. Therefore, the divergence rate of the KL divergence when dealing with NSCE processes is equal to:

$$\Delta KL^{(1,2)} = \frac{1}{2}\Big[\sum_{m=1}^{M_1} \frac{\gamma_{m,1}}{\sigma_2^2}(1 - \sum_{n=1}^{M_2} \delta_{\theta_{m,1},\theta_{n,2}}) + \frac{\sigma_1^2}{\sigma_2^2} - 1 - \ln\frac{\sigma_1^2}{\sigma_2^2}\Big] \tag{C.6}$$

Appendix C.2. Link with the Rényi divergence rate

Let us recall the expression of the increment of the Rényi divergence we obtained in (67):

$$\begin{aligned}
\Delta RD_k^{(1,2)}(\alpha) &= -\frac{1}{2(\alpha-1)} \ln\Big(\frac{|Q_{k,2}|^\alpha}{|Q_{k+1,2}|^\alpha} \frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|} \frac{|Q_{k,1}|^{1-\alpha}}{|Q_{k+1,1}|^{1-\alpha}}\Big) \\
&+ \frac{\alpha}{2}\Big(\mathrm{Tr}\big(Q_{k+1,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_{k+1}\Delta\mu_{k+1}^T\big) - \mathrm{Tr}\big(Q_{k,wei}^{-1}(1-\alpha,\alpha)\Delta\mu_k\Delta\mu_k^T\big)\Big) \\
&= \frac{E(\alpha)}{D(\alpha)}
\end{aligned}$$

where:

$$\begin{aligned}
E_k(\alpha) &= \ln\Big(\frac{|Q_{k,2}|^\alpha}{|Q_{k+1,2}|^\alpha} \frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|} \frac{|Q_{k,1}|^{1-\alpha}}{|Q_{k+1,1}|^{1-\alpha}}\Big) \\
&= \ln\Big(\frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|}\Big) + \alpha\ln\Big(\frac{|Q_{k,2}|}{|Q_{k+1,2}|}\Big) + (1-\alpha)\ln\Big(\frac{|Q_{k,1}|}{|Q_{k+1,1}|}\Big)
\end{aligned} \tag{C.7}$$

and

$$D(\alpha) = -2(\alpha-1) \tag{C.8}$$

As $Q_{k,wei}(1-\alpha,\alpha)$ and $Q_{k+1,wei}(1-\alpha,\alpha)$ respectively tend to $Q_{k,2}$ and $Q_{k+1,2}$ when α tends to 1, $E(\alpha)$ tends to 0 when α tends to 1 while $D(\alpha)$ also tends to 0. Therefore, one has to find another solution to deduce the limit of the first part of the increment. The idea is to use L'Hôspital Rule, *i.e.* to study the limit of the ratio of the derivates $E'(\alpha)$ and $D'(\alpha)$ when α tends to 1. Given (C.8), $D'(\alpha) = -2$. Let us now look at the derivate of $E_k(\alpha)$:

$$E_k'(\alpha) = \underbrace{\frac{d}{d\alpha}(\ln\Big(\frac{|Q_{k+1,wei}(1-\alpha,\alpha)|}{|Q_{k,wei}(1-\alpha,\alpha)|}\Big))}_{E_{k,1}'(\alpha)} + \underbrace{\ln\Big(\frac{|Q_{k,2}|}{|Q_{k+1,2}|}\Big) - \ln\Big(\frac{|Q_{k,1}|}{|Q_{k+1,1}|}\Big)}_{E_{k,2}'(\alpha)} \tag{C.9}$$

According to (19), $E'_{k,2}(\alpha)$ does not depend on α. It satisfies:

$$\lim_{k\to+\infty} E'_{k,2}(\alpha) = \lim_{k\to+\infty} \ln\Big(\frac{|Q_{k,2}|}{|Q_{k+1,2}|}\Big) - \ln\Big(\frac{|Q_{k,1}|}{|Q_{k+1,1}|}\Big) = \ln\frac{\sigma_1^2}{\sigma_2^2} \tag{C.10}$$

Let us focus our attention on $E'_{k,1}(\alpha)$ which can be rewritten according to (10) and by adjusting the notations as follows:

$$E'_{k,1}(\alpha) = \frac{d}{d\alpha}\Big(\ln(r_{0,wei} - R_{k,wei}^H Q_{k,wei}^{-1}(1-\alpha,\alpha)R_{k,wei})\Big) \tag{C.11}$$

with

$$\begin{cases} r_{0,wei} & = (1-\alpha)r_{0,1} + \alpha r_{0,2} \\ & = (1-\alpha)(Tr(\phi_1)+\sigma_1^2) + \alpha(Tr(\phi_2)+\sigma_2^2) \\ Q_{k,wei}(1-\alpha,\alpha) & = (1-\alpha)Q_{k,1} + \alpha Q_{k,2} \\ & = (1-\alpha)(S_{k,1}\Phi_1 S_{k,1}^H + \sigma_1^2 I_k) + \alpha(S_{k,2}\Phi_2 S_{k,2}^H + \sigma_2^2 I_k) \\ R_{k,wei} & = \begin{bmatrix} I_k & 0_{k,1} \end{bmatrix} Q_{k+1,wei}(1-\alpha,\alpha) \begin{bmatrix} 0_{k,1} \\ 1 \end{bmatrix} \\ & = (1-\alpha)S_{k,1}\Phi_1 \begin{bmatrix} e^{jk\theta_{1,1}} & \dots & e^{jk\theta_{1,M_1}} \end{bmatrix}^H \\ & +\alpha S_{k,2}\Phi_2 \begin{bmatrix} e^{jk\theta_{2,1}} & \dots & e^{jk\theta_{2,M_2}} \end{bmatrix}^H \\ & = (1-\alpha)S_{k,1}\Phi_1\Psi_{k,1}^H + \alpha S_{k,2}\Phi_2\Psi_{k,2}^H \end{cases} \tag{C.12}$$

In the following, our purpose is to express $E'_{k,1}(\alpha)$ and look at its limit when k tends to infinity and α tends to 1. Therefore, one has:

$$E'_{k,1}(\alpha) = \frac{1}{r_{0,wei} - R_{k,wei}^H Q_{k,wei}^{-1}(1-\alpha,\alpha)R_{k,wei}} \times \tag{C.13}$$
$$\Big(-r_{0,1} + r_{0,2} - R_{k,wei}^H Q_{k,wei}^{-1}(1-\alpha,\alpha)Q_{k,wei}^{-1}(1-\alpha,\alpha)(-Q_{k,1}+Q_{k,2})R_{k,wei}\Big)$$

In (C.13), the limit of $r_{0,wei} - R_{k,wei}^H Q_{k,wei}^{-1}(1-\alpha,\alpha)R_{k,wei}$ when k tends to infinity is $(1-\alpha)\sigma_1^2 + \alpha\sigma_2^2$. Consequently, the limit when α tends to 1 is σ_2^2. In addition, $r_{0,2} - r_{0,1}$ does not depend on k and α and hence remains unchanged when taking the limits. Therefore, one has to study the limit of $-R_{k,wei}^H Q_{k,wei}^{-1}(1-\alpha,\alpha)Q_{k,wei}^{-1}(1-\alpha,\alpha)(-Q_{k,1}+Q_{k,2})R_{k,wei}$ when k increases and tends to infinity and α tends to 1. To this end, let us consider (12) and adjust its expression to the current case. Introducing $\sigma_{wei}^2 = (1-\alpha)\sigma_1^2 + \alpha\sigma_2^2$,

Terms to be studied	$\lim_{\alpha\to 1} L_i$
$L_1 = \lim_{k\to+\infty} \frac{1}{\sigma_{wei}^4} R_{k,wei}^H Q_{k,1} R_{k,wei}$	$\lim_{k\to+\infty} \frac{k^2}{\sigma_2^4}\Psi_{k,2}\Phi_2\Lambda_{2,1}\Phi_1\Lambda_{2,1}^H\Phi_2\Psi_{k,2}^H + \frac{\sigma_1^2 k}{\sigma_2^2}\Psi_{k,2}\Phi_2^2\Psi_{k,2}^H$
$L_2 = \lim_{k\to+\infty} -\frac{1}{\sigma_{wei}^4} R_{k,wei}^H Q_{k,2} R_{k,wei}$	$-\lim_{k\to+\infty} \frac{k^2}{\sigma_2^4}\Psi_{k,2}\Phi_2^3\Psi_{k,2}^H + \frac{k}{\sigma_2^2}\Psi_{k,2}\Phi_2^2\Psi_{k,2}^H$
$L_3 = -\lim_{k\to+\infty} \frac{1}{k\sigma_{wei}^4} R_{k,wei}^H S_{k,wei} S_{k,wei}^H Q_{k,1} R_{k,wei}$	$\lim_{k\to+\infty} -\frac{k^2}{\sigma_2^4}\Psi_{k,2}\Phi_2\Lambda_{2,1}\Phi_1\Lambda_{2,1}^H\Phi_2\Psi_{k,2}^H - \frac{\sigma_1^2 k}{\sigma_2^4}\Psi_{k,2}\Phi_2^2\Psi_{k,2}^H$
	$= -\lim_{\alpha\to 1} L_1$
$L_4 = \lim_{k\to+\infty} -\frac{1}{k\sigma_{wei}^4} R_{k,wei}^H S_{k,wei} S_{k,wei}^H Q_{k,2} R_{k,wei}$	$\lim_{k\to+\infty} -\frac{k^2}{\sigma_2^4}\Psi_{k,2}\Phi_2^3\Psi_{k,2}^H - \frac{k}{\sigma_2^2}\Psi_{k,2}\Phi_2^2\Psi_{k,2}^H$
	$= -\lim_{\alpha\to 1} L_2$
$L_5 = \lim_{k\to+\infty} \frac{R_{k,wei}^H S_{k,wei}\Phi_{wei}^{-2} S_{k,wei}^H Q_{k,1} R_{k,wei}}{k^3}$	$\sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\gamma_{m,1}\delta_{\theta_{1,m},\theta_{2,n}}$
$L_6 = \lim_{k\to+\infty} \frac{R_{k,wei}^H S_{k,wei}\Phi_{wei}^{-2} S_{k,wei}^H Q_{k,2} R_{k,wei}}{k^3}$	$\lim_{k\to+\infty}\Psi_{k,2}\Phi_2\Psi_{k,2}^H + \sigma_2^2 = Tr(\Phi_2) = r_{0,2} - \sigma_2^2$

Table C.2: Terms to be studied and their equivalent expressions when α tends to 1 using the asymptotic properties

one has:

$$\lim_{k\to+\infty} Q_{k,wei}^{-1}(1-\alpha,\alpha) \tag{C.14}$$
$$= \lim_{k\to+\infty}\left(\frac{1}{\sigma_{wei}^2} I_k - \frac{1}{k\sigma_{wei}^2} S_{k,wei} S_{k,wei}^H + \frac{1}{k^2} S_{k,wei}(\Phi_{wei} - \frac{\sigma_{wei}^2}{k})^{-1} S_{k,wei}^H\right)$$

Therefore taking advantage of the asymptotic properties of $S_{k,wei}$ similar to those of $S_{k,i}$ for $i = 1, 2$, one has after development and simplification:

$$\lim_{k\to+\infty} Q_{k,wei}^{-1}(1-\alpha,\alpha) Q_{k,wei}^{-1}(1-\alpha,\alpha) \tag{C.15}$$
$$= \lim_{k\to+\infty}\left(\frac{1}{\sigma_{wei}^4} I_k - \frac{1}{k\sigma_{wei}^4} S_{k,wei} S_{k,wei}^H + \frac{1}{k^3} S_{k,wei}\Phi_{wei}^{-2} S_{k,wei}^H\right)$$

The next step is to pre-multiply the above expression by $-R_{k,wei}^H$ and post-multiply it by $(Q_{k,2} - Q_{k,1})R_{k,wei}$. By introducing $\Lambda_{2,1} = \lim_{k\to+\infty} \frac{1}{k} S_{k,2}^H S_{k,1}$ which is a matrix of dimension $M_2 \times M_1$ whose elements are equal to $\delta_{\theta_{i,m},\theta_{i',n}}$ which denotes the Kronecker delta.[14], this amounts to looking at six limits. After mathematical development and simplification one gets the following results summarized in the table C.1.

Using the result given in table C.1, the limit of $E'_{k,1}(\alpha)$ when α tends to 1

[14] In other words, this means that it is equal to 1 when when the normalized angular frequency of n^{th} component of $S_{k,2}$ is equal to the m^{th} component of $S_{k,1}$ and zero otherwise

verifies:

$$\lim_{\alpha\to 1} E'_{k,1}(\alpha) = \frac{1}{\sigma_2^2}\left(-r_{0,1} + r_{0,2} + \sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\gamma_{m,1}\delta_{\theta_{1,m},\theta_{2,n}} - r_{0,2} + \sigma_2^2\right) \tag{C.16}$$
$$= -\frac{1}{\sigma_2^2}\left(\sum_{m=1}^{M_1}\gamma_{m,1} + \sigma_1^2 - \sum_{m=1}^{M_1}\sum_{n=1}^{M_2}\gamma_{m,1}\delta_{\theta_{1,m},\theta_{2,n}}\right) + 1$$
$$= 1 - \frac{1}{\sigma_2^2}\sum_{m=1}^{M_1}\gamma_{m,1}\left(1 - \sum_{n=1}^{M_2}\delta_{\theta_{1,m},\theta_{2,n}}\right) - \frac{\sigma_1^2}{\sigma_2^2}$$

Consequently, the limit of $E'_k(\alpha)$ is equal to:

$$\lim_{\alpha\to 1} E'_k(\alpha) = 1 - \frac{1}{\sigma_2^2}\sum_{m=1}^{M_1}\gamma_{m,1}\left(1 - \sum_{n=1}^{M_2}\delta_{\theta_{1,m},\theta_{2,n}}\right) - \frac{\sigma_1^2}{\sigma_2^2} + \ln\frac{\sigma_1^2}{\sigma_2^2} \tag{C.17}$$

Finally, one obtains:

$$\lim_{\alpha\to 1}\frac{E'_k(\alpha)}{D'(\alpha)} = -\frac{1}{2} + \frac{1}{2\sigma_2^2}\sum_{m=1}^{M_1}\gamma_{m,1}\left(1 - \sum_{n=1}^{M_2}\delta_{\theta_{1,m},\theta_{2,n}}\right) + \frac{\sigma_1^2}{2\sigma_2^2} - \frac{1}{2}\ln\frac{\sigma_1^2}{\sigma_2^2} \tag{C.18}$$

The divergence rate of the KL divergence is hence retrieved for NSCE processes.

Appendix C.3. Link between the beta-divergence and the KL divergence in the Gaussian case

Let us recall the expression of the β-divergence given in (62) and rewrite it a bit differently:

$$BeD_k^{(1,2)}(\beta) = \frac{1}{\beta^{\frac{k}{2}+1}(2\pi)^{\frac{k(\beta-1)}{2}}}\left(\frac{1}{(\beta-1)|Q_{k,1}|^{\frac{(\beta-1)}{2}}} + \frac{1}{|Q_{k,2}|^{\frac{(\beta-1)}{2}}}\right) \tag{C.19}$$
$$-\frac{1}{\beta-1}\frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}\exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)$$

or equivalently:

$$BeD_k^{(1,2)}(\beta) = \frac{1}{\beta^{\frac{k}{2}+1}(2\pi)^{\frac{k(\beta-1)}{2}}}\Big(-\frac{1}{|Q_{k,1}|^{\frac{(\beta-1)}{2}}} \tag{C.20}$$
$$+\frac{1}{|Q_{k,2}|^{\frac{(\beta-1)}{2}}}\Big) + \frac{1}{(\beta-1)(2\pi)^{\frac{k(\beta-1)}{2}}}\frac{1}{\beta^{\frac{k}{2}+1}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}$$
$$-\frac{1}{\beta-1}\frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}}|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}}\exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)$$

In the above expression, the first term tends to 0 when β tends to 1. Let us now look at the sum of the last two terms:

$$\begin{aligned} F_k(\beta) &= \frac{1}{(\beta-1)(2\pi)^{\frac{k(\beta-1)}{2}}} \times \Big(\frac{1}{\beta^{\frac{k}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}} \\ &- \frac{1}{|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}} \exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)\Big) \\ &= \frac{1}{J_k(\beta)} \times H_k(\beta) \end{aligned} \tag{C.21}$$

where:

$$\lim_{\beta\to 1} J_k(\beta) = \lim_{\beta\to 1}(\beta-1)(2\pi)^{\frac{k(\beta-1)}{2}} = 0 \tag{C.22}$$

and

$$\begin{aligned} \lim_{\beta\to 1} H_k(\beta) &= \lim_{\beta\to 1}\Big(\frac{1}{\beta^{\frac{k}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}} \\ &- \frac{1}{|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}} \exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)\Big) = 0 \end{aligned} \tag{C.23}$$

Therefore, we suggest using L'Hôspital rule once again. One has after expressing $(2\pi)^{\frac{k(\beta-1)}{2}}$ as $\exp\left(\frac{k(\beta-1)}{2}\ln(2\pi)\right)$:

$$\lim_{\beta\to 1} J_k'(\beta) = \lim_{\beta\to 1}(2\pi)^{\frac{k(\beta-1)}{2}} + (\beta-1)\frac{k}{2}(2\pi)^{\frac{k(\beta-1)}{2}} = 1 \tag{C.24}$$

Let us now look at $\lim_{\beta\to 1} H_k'(\beta)$. To this end, one must express the derivate of $h_1(\beta) = \frac{1}{\beta^{\frac{k}{2}}|Q_{k,1}|^{\frac{(\beta-1)}{2}}}$ and then the derivate of the following quantity: $h_2(\beta) = -\frac{1}{|Q_{k,wei}(\beta-1,1)|^{\frac{1}{2}}|Q_{k,2}|^{\frac{\beta-2}{2}}} \exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right)$:

$$\begin{aligned} \lim_{\beta\to 1} h_1'(\beta) &= \lim_{\beta\to 1} \frac{d\left(\beta^{-\frac{k}{2}}|Q_{k,1}|^{\frac{(1-\beta)}{2}}\right)}{d\beta} \\ &= \lim_{\beta\to 1} -\frac{k}{2}\beta^{-\frac{k}{2}-1}|Q_{k,1}|^{\frac{(1-\beta)}{2}} + \beta^{-\frac{k}{2}}|Q_{k,1}|^{\frac{(1-\beta)}{2}} \times \left(-\frac{1}{2}\ln|Q_{k,1}|\right) \\ &= -\frac{k}{2} - \frac{1}{2}\ln|Q_{k,1}| \end{aligned} \tag{C.25}$$

Let us now decompose $h_2(\beta)$ as a product of three terms all depending on β. The first one $h_{21}(\beta)$ is given by:

$$\begin{aligned} h_{21}(\beta) &= -|Q_{k,wei}(\beta-1,1)|^{-\frac{1}{2}} = -|(\beta-1)Q_{k,1} + Q_{k,2}|^{-\frac{1}{2}} \\ &= -|(\beta-1)Q_{k,2}^{-1}Q_{k,1} + I_k|^{-\frac{1}{2}}|Q_{k,2}|^{-\frac{1}{2}} \end{aligned} \tag{C.26}$$

with:

$$\lim_{\beta \to 1} h_{21}(\beta) = -|Q_{k,2}|^{-\frac{1}{2}} \tag{C.27}$$

Then, $Q_{k,2}^{-1}Q_{k,1}$ can be expressed from the eigenvalue decomposition with the eigenvalues $\{\lambda_i\}_{i=1,\dots,k}$. Consequently, (C.26) can be rewritten this way:

$$h_{21}(\beta) = -\left(\prod_{i=1}^{k}((\beta-1)\lambda_i+1)\right)^{-\frac{1}{2}}|Q_{k,2}|^{-\frac{1}{2}} \tag{C.28}$$

Therefore, the limit of the derivate $h'_{21}(\beta)$ when β tends to 1 is equal to:

$$\lim_{\beta \to 1} h'_{21}(\beta) = -|Q_{k,2}|^{-\frac{1}{2}} \lim_{\beta \to 1}\left(\prod_{i=1}^{k}((\beta-1)\lambda_i+1)\right)^{-\frac{1}{2}} \times \left(-\frac{1}{2}\sum_{i=1}^{k}\frac{\lambda_i}{1+(\beta-1)\lambda_i}\right) \tag{C.29}$$

$$= \frac{|Q_{k,2}|^{-\frac{1}{2}}}{2}\sum_{i=1}^{k}\lambda_i = \frac{|Q_{k,2}|^{-\frac{1}{2}}}{2}Tr(Q_{k,2}^{-1}Q_{k,1})$$

Let us now look at the second term $h_{22}(\beta) = |Q_{k,2}|^{\frac{2-\beta}{2}}$ and the limit of its derivate when β tends to 1. One can first notice:

$$\lim_{\beta \to 1} h_{22}(\beta) = |Q_{k,2}|^{\frac{1}{2}} \tag{C.30}$$

In addition, its derivate satisfies:

$$\lim_{\beta \to 1} h'_{22}(\beta) = \lim_{\beta \to 1}|Q_{k,2}|^{\frac{2-\beta}{2}} \times \frac{1}{2}\ln|Q_{k,2}| = \frac{|Q_{k,2}|^{\frac{1}{2}}}{2}\ln|Q_{k,2}| \tag{C.31}$$

Let us finally look at the third term defined as follows:

$$h_{23}(\beta) = \exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right) \tag{C.32}$$

with:

$$\lim_{\beta \to 1} h_{23}(\beta) = 1 \tag{C.33}$$

Its derivate satisfies:

$$\lim_{\beta \to 1} h'_{23}(\beta) = \lim_{\beta \to 1}\exp\left(-\frac{(\beta-1)}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k\right) \tag{C.34}$$

$$\times\left(-\frac{1}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k - \frac{(\beta-1)}{2}\Delta\mu_k^T\frac{dQ_{k,wei}^{-1}}{d\beta}(\beta-1,1)\Delta\mu_k\right)$$

$$= -\frac{1}{2}\Delta\mu_k^T Q_{k,wei}^{-1}(\beta-1,1)\Delta\mu_k = -\frac{1}{2}Tr\left(Q_{k,2}^{-1}\Delta\mu_k\Delta\mu_k^T\right)$$

To end up the reasoning, one just has to express $\lim_{\beta\to 1} h_2'(\beta) = \lim_{\beta\to 1} h_{21}' h_{22} h_{23} + h_{21} h_{22}' h_{23} + h_{21} h_{22} h_{23}'$. One obtains:

$$\begin{aligned}\lim_{\beta\to 1} h_2'(\beta) &= \frac{|Q_{k,2}|^{-\frac{1}{2}}}{2} Tr(Q_{k,2}^{-1} Q_{k,1}) |Q_{k,2}|^{\frac{1}{2}} \\ &- |Q_{k,2}|^{-\frac{1}{2}} \frac{|Q_{k,2}|^{\frac{1}{2}}}{2} \ln |Q_{k,2}| + |Q_{k,2}|^{-\frac{1}{2}} |Q_{k,2}|^{\frac{1}{2}} \frac{1}{2} Tr\left(Q_{k,2}^{-1} \Delta\mu_k \Delta\mu_k^T\right) \\ &= \frac{1}{2} Tr(Q_{k,2}^{-1} Q_{k,1}) + \frac{1}{2} \ln |Q_{k,2}| + \frac{1}{2} Tr\left(Q_{k,2}^{-1} \Delta\mu_k \Delta\mu_k^T\right)\end{aligned} \tag{C.35}$$

Finally, as $\lim_{\beta\to 1} H_k'(\beta) = \lim_{\beta\to 1} h_1'(\beta) + h_2'(\beta)$, one can deduce that:

$$\lim_{\beta\to 1} H_k'(\beta) = \frac{1}{2}\left(-k + Tr(Q_{k,2}^{-1} Q_{k,1}) - \ln \frac{|Q_{k,1}|}{|Q_{k,2}|} + Tr\left(Q_{k,2}^{-1} \Delta\mu_k \Delta\mu_k^T\right)\right) \tag{C.36}$$

and finally:

$$\begin{aligned}\lim_{\beta\to 1} \frac{H_k'(\beta)}{J_k'(\beta)} &= \frac{1}{2}\left(-k + Tr(Q_{k,2}^{-1} Q_{k,1}) - \ln \frac{|Q_{k,1}|}{|Q_{k,2}|} + Tr\left(Q_{k,2}^{-1} \Delta\mu_k \Delta\mu_k^T\right)\right) \\ &= \Delta KL^{(1,2)}\end{aligned} \tag{C.37}$$

References

[1] S. Amari and H. Nagaoka. Methods of Information Geometry. Oxford University Press: New York,NY, USA, 2000.

[2] M. Basseville. Divergence measures for statistical data processing. an annotated bibliography. Signal Processing, 93 (4):621–633, 2013.

[3] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of the Calcutta Mathematical Society, 35:99–109, 1943.

[4] L. Bombrun, N. E. Lasmar, Y. Berthoumieu, and G. Verdoolaege. Multivariate texture retrieval using the SIRV representation and the geodesic distance. IEEE ICASSP, pages 865–868, 2011.

[5] H. Chernoff. Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Annals of Mathematical Statistics, 23:493–507, 1952.

[6] A. Cichocki and S.-I. Amari. Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. Entropy, 12:1532–1568, 2010.

[7] G. Ciuperca, V. Girardin, and L. Lhote. Computation and estimation of generalized en-tropy rates for denumerable markov chains. IEEE Transactions on Information Theory, 57:4026–4034, 2011.

[8] J. E. Contreras-Reyes. Analyzing fish condition factor index through skew-gaussian information theory quantifiers. Fluctuation and Noise Letters, 15 (2):1–16, 2016.

[9] J. E. Contreras-Reyes and R. B. Arellano-Valle. Kullback–Leibler divergence measure for multivariate skew-normal distributions. Entropy, 14 (9):1606–1626, 2012.

[10] D. C. de Souza, R. F. Vigelis, and C. C. Cavalcante. Geometry induced by a generalization of Rényi divergence. Entropy, 18, 407, 2015.

[11] H. Fujisawa and S. Eguchi. Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis, 99:2053–2081, 2008.

[12] T. T. Georgiou and A. Lindquist. A convex optimization approach to ARMA modeling. IEEE Transactions on Automatic Control, 53:1108–1119, 2008.

[13] M. Gil. On Rényi divergence measures for continuous alphabet sources. PhD Thesis, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, 2011.

[14] M. Gil, F. Alajaji, and T. Linder. Rényi divergence measures for commonly used univariate continuous distributions. Information Sciences, 249:124–131, 2013.

[15] V. Girardin and P. Lhote, L. Regnault. Different closed-form expressions for generalized entropy rates of markov chains. Methodology and Computing in Applied Probability, 21:1431–1452, 2019.

[16] J. Gomez-Garde and V. Latora. Entropy rate of diffusion processes on complex networks. Physical Review E, 78:065102, 2009.

[17] E. Grivel, R. Diversi, and F. Merchan. Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. Elsevier Digital Signal Processing, 116:103089, 2021.

[18] E. Grivel, M. Saleh, and S.-M. Omar. Jeffrey's divergence between complex-valued sinusoidal processes. EUSIPCO, 2017.

[19] E. Grivel, M. Saleh, and S.-M. Omar. Interpreting the asymptotic increment of Jeffrey's divergence between some random processes. Elsevier Digital Signal Processing, 75, (4):120–133, 2018.

[20] S. Ihara. Information Theory for Continuous Systems. World Scientific, 1993.

[21] H. Jeffreys. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, 186:453–461, September 1946.

[22] P. Kluza. On Jensen-Rényi and Jeffreys-Rényi type f-divergences induced by convex functions. Physica A: Statistical Mechanics and its Applications, 548:122527, 2020.

[23] P. Kluza and M. Niezgoda. Generalizations of Crooks and Lin's results on Jeffreys-Csiszár and Jensen-Csiszár f-divergences. Physica A: Statistical Mechanics and its Applications, 463:383–393, 2016.

[24] S. Kullback and R. A. Leibler. On Information and Sufficiency. The Annals of Mathematical Statistics, 22 (1):79–86, 1951.

[25] M. Lasta and R. Shumway. Detecting abrupt changes in a piecewise locally stationary time series. Journal of Multivariate Analysis, 99:191–214, 2008.

[26] J. Lin. Divergence measures based on the shannon entropy. IEEE Transactions on Information Theory, 37:145–151, 1991.

[27] C. Magnant, E. Grivel, A. Giremus, B. Joseph, and L. Ratton. Jeffrey's divergence for state-space model comparison. Signal Processing, 114:61–74, September 2015.

[28] R. Murthy, I. Pavlidis, and P. Tsiamyrtzis. Touchless monitoring of breathing function. IEEE EMBS, pages 1196–1199, 2004.

[29] M. Najim. Modeling, estimation and optimal filtering in signal processing. Wiley, 2010.

[30] X. Nguyen, M. J. Wainwright, and M. I. Jordan. Estimating divergence functionals and the likelihood ratio by convex risk minimization. IEEE Transactions on Information Theory, 56, n°11:5847–5861, 2010.

[31] F. Nielsen and R. Nock. Statistical exponential families: a digest with slash cards. Computing Research Repository (CoRR) - arXiv (CoRR abs/0911.4863), 11, 2009.

[32] F. Nielsen and R. Nock. A closed-form expression for the Sharma-Mittal entropy of exponential families. Journal of Physics A: Mathematical and Theoretical, 45, 2012.

[33] F. Nielsen and R. Nock. On Rényi and Tsallis entropies and divergences for exponential families. Journal of Physics A: Mathematical and Theoretical, 45, (3), 2012.

[34] F. Nielsen and R. Nock. On the chi square and higher-order chi distances for approximating f-divergences. IEEE Signal Processing Letters, 21, (1):10–13, 2014.

[35] Z. Rached, F. Alajaji, and L. L. Campbell. Rényi's entropy rate for discrete markov sources. Conference on Information Sciences and Systems (CISS), pages 613–618, 1999.

[36] Z. Rached, F. Alajaji, and L. L. Campbell. Rényi's divergence and entropy rates for finite alphabet markov sources. IEEE Transactions on Information Theory, 47:1553–1561, 2001.

[37] Z. Rached, F. Alajaji, and L. L. Campbell. The Kullback–Leibler divergence rate between markov sources. IEEE Transactions on Information Theory, 50:917–921, 2004.

[38] P. Regnault, V. Girardin, and L. Lhote. Weighted closed form expressions based on escort distributions for Rényi entropy rates of markov chains. GSI'17, 2017.

[39] A. Rényi. On measures of entropy and information. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Bulletin of the Calcutta Mathematical Society, 1:547–561, 1961.

[40] P.R. Scalassara, M. E. Dajer, J. L. Marrara, C. D. Maciel, and J. C. Pereira. Analysis of voice pathology evolution using entropy rate. IEEE International Symposium on Multimedia, pages 580–585, 2008.

[41] A. Schutz, L.Bombrun, Y. Berthoumieu, and M. Najim. Centroid-based texture classification using the generalized gamma distribution. EUSIPCO, pages 1–5, 2013.

[42] M. Sugiyama, S. Liu, M. Christoffel du Plessis, M. Yamanaka, T. Suzuki, and T. Kanamori. Direct divergence approximation between probability distributions and its applications in machine learning. Journal of Computing Science and Engineering, 7, no. 2:99–111, 2013.

[43] M. Sugiyama, T. Suzuki, S. Nakajima, P. von Bunau, and M. Kawanabe. Direct importance estimation for covariate shift adapation. Annals of the Institute, 60, n°4:699–746, 2008.

[44] R. F. Vigelis, L. H.F. de Andrade, and C. C. Cavalcante. Conditions for the existence of a generalization of Rényi divergence. Physica A, 558:124953, 2020.

[45] M. Yamada and M. Sugiyama. Direct importance estimation with gaussian mixture models. IEICE Transactions on Information and Systems, E92-D, n°10:2159–2162, 2009.