

Taux de divergence ou d'entropie croisée pour la comparaison de processus aléatoires à mémoire courte ou à mémoire longue

Eric GRIVEL¹

¹Laboratoire IMS
UMR CNRS 5218, 351 cours de la libération, 33400 Talence, France
eric.grivel@ims-bordeaux.fr

Résumé – Dans cette communication, notre objectif est d'étudier certaines propriétés de mesures telles que les divergences ou les entropies croisées de processus stochastiques. Plus particulièrement, nous proposons d'exprimer ce que l'on appelle le taux d'entropie croisée de deux bruits blancs à intégration fractionnaire Gaussiens. Ensuite, nous nous intéressons au taux de β -divergence de deux processus autorégressifs à moyenne ajustée (ARMA) Gaussiens. Selon les cas, ces taux peuvent être utilisés pour comparer ces signaux.

Abstract – In this paper our purpose is to study some properties of measures such as divergences or cross-entropies of stochastic processes. More particularly, we propose to derive what is called the rate of the cross-entropy between two Gaussian fractionally integrated white noises. Then, we focus our attention on the rate of the β -divergence between two Gaussian autoregressive with moving average (ARMA) processes. Depending on the case, these rates can be used to compare signals.

1 Introduction

Dans de nombreuses applications, notamment liées à la parole ou au biomédical, nous sommes amenés à comparer des signaux aléatoires. Pour ce faire, plusieurs approches sont envisageables : certaines consistent à modéliser les signaux, à estimer les paramètres des modèles choisis, puis à les comparer. D'autres reposent sur des distances spectrales comme la distance log-spectrale. Une autre manière d'opérer est de partir des densités de probabilité. Ainsi, la divergence de Kullback-Leibler (KL) permet de rendre compte de la dissimilarité entre deux densités de probabilité. Si elle est souvent utilisée dans nos disciplines, d'autres divergences ont été proposées. On peut citer celles qui sont une fonction du coefficient de Chernoff comme les divergences de Chernoff, de Bhattacharyya, de Rényi et de Sharma-Mittal ou encore les α -divergences. Certaines sont des f -divergences, mais elles se déduisent aussi d' α -divergences, c'est le cas de la divergence de Pearson. Quant aux divergences de Bregman, elles incluent les β -divergences.

De nombreux travaux sur les divergences et plus généralement sur des mesures utilisées en théorie de l'information comme l'entropie et l'entropie croisée ont été menés récemment [9, 12, 7]. Plusieurs auteurs se sont aussi intéressés à l'expression de ces grandeurs par unité de temps. Ainsi, le taux d'entropie d'une source ou d'un processus stochastique (c'est-à-dire dans le cas "classique" la limite quand elle existe de l'entropie jointe de n variables divisée par n quand n tend vers l'infini. A noter que le taux peut être défini d'une autre manière en fonction du type de mesure) a été traité par des auteurs comme Rached, Ciuperca, Girardin, Lhote, Regnault et Limnios dans plusieurs contextes : estimation du taux des entropies de Shannon et de Rényi ou encore celui des (h, Φ) -entropies

pour des processus sans mémoire ou des chaînes de Markov, des processus semi-markoviens à temps discret ou continu [10] [1] [11] [3]. De plus, dans [2], des expressions du taux des divergences de KL et de Rényi pour des processus à moyenne nulle Gaussiens ont été obtenues sous la forme d'intégrales faisant intervenir les densités spectrales de puissance (DSP) des processus considérés. Dans [5], nous avons exploité les propriétés des processus ARMA pour déduire les taux de divergence et avons fait le lien entre le taux de divergence de KL et la distance d'Itakura-Saito. Dans cette communication, nous étudions si ces taux sont toujours adaptés pour comparer des processus aléatoires. Dans un premier temps, nous analysons le taux d'entropie croisée de bruits blancs à intégration fractionnaire (BBIF) Gaussiens. Puis, pour compléter le travail présenté dans [6], le taux de β -divergence de processus ARMA Gaussiens est abordé. Le reste de cette communication s'organise comme suit : des rappels sur les processus et leurs propriétés sont faits dans la section 2. Puis, les deux cas proposés sont traités et illustrés dans la partie 3. Enfin, des perspectives sont données. Dans la suite, $\Gamma(\cdot)$ est définie comme suit :

$$\Gamma(k) = \begin{cases} \int_0^{+\infty} t^{k-1} e^{-t} dt & \text{for } k > 0 \\ k^{-1} \Gamma(k+1) & \text{for } k < 0 \\ +\infty & k = 0 \end{cases} \quad (1)$$

Le théorème binomial se formule ainsi :

$$(1 - z^{-1})^g = \sum_{k=0}^{+\infty} \frac{\Gamma(k-g)}{\Gamma(-g)\Gamma(k+1)} z^{-k} \quad (2)$$

De plus, les notations suivantes sont introduites pour la définition des divergences et des entropies :

$$\left\{ \begin{aligned} I_1(\mathbf{p}_1, \mathbf{p}_2, a, b) &= \int_{X_k} \mathbf{p}_1^a(X_k) \mathbf{p}_2^b(X_k) dX_k \end{aligned} \right. \quad (3a)$$

$$\left\{ \begin{aligned} I_2(\mathbf{p}_1, a) &= \int_{X_k} \mathbf{p}_1^a(X_k) dX_k \end{aligned} \right. \quad (3b)$$

Si p_1 et p_2 sont deux densités de probabilité Gaussiennes réelles de matrice de covariance $Q_{k,1}$ et $Q_{k,2}$ de taille $k \times k$ et dont la différence des vecteurs moyennes vaut $\Delta\mu_k$, il vient en posant $Q_{k,b,a} = bQ_{k,1} + aQ_{k,2}$, avec $a > 0$ et $b > 0$, et $|\cdot|$ le déterminant :

$$\begin{cases} I_1(p_1, p_2, a, b) = \frac{|Q_{k,1}|^{\frac{1-a}{2}} |Q_{k,2}|^{\frac{1-b}{2}}}{(2\pi)^{\frac{k(a+b-1)}{2}} |Q_{k,b,a}|^{\frac{1}{2}}} \exp\left(-\frac{ab}{2} \Delta\mu_k^T Q_{k,b,a}^{-1} \Delta\mu_k\right) \\ I_2(p_1, a) = \frac{1}{a^{\frac{k}{2}} (2\pi)^{\frac{k(a-1)}{2}} |Q_{k,1}|^{\frac{1}{2}}} \end{cases}$$

Dans la suite, nous rappelons les propriétés des signaux pour lesquels nous allons déduire des taux de différentes mesures. L'indice i sera utilisé pour faire référence au $i^{\text{ème}}$ processus.

2 Processus étudiés

2.1 Processus ARMA

Soit un processus ARMA d'ordre (p_i, q_i) dont le $t^{\text{ème}}$ échantillon $x_{t,i}$ vérifie :

$$x_{t,i} = - \sum_{j=1}^{p_i} a_{j,i} x_{t-j,i} + \sum_{j=0}^{q_i} b_{j,i} u_{t-j,i} \text{ avec } i = 1, 2 \quad (4)$$

où $\{a_{j,i}\}_{j=1,\dots,p_i}$, $b_{0,i} = 1$ et $\{b_{j,i}\}_{j=1,\dots,q_i}$ sont les paramètres ARMA alors que le processus générateur $u_{t,i}$ est blanc Gaussien de moyenne nulle et de variance $\sigma_{u,i}^2$. Ce processus est stationnaire au sens large et à mémoire courte puisque sa fonction de covariance est absolument sommable. Il peut aussi s'interpréter tel la sortie d'un filtre linéaire causal à réponse impulsionnelle infinie dont l'entrée est $u_{t,i}$. La fonction de transfert associée est définie à partir des pôles de module strictement inférieur à 1 pour garantir la stabilité, et des zéros $\{z_{l,i}\}_{l=1,\dots,q_i}$. Quand ces derniers ont un module strictement inférieur à 1, on parle de processus ARMA à minimum de phase. Si ce n'est pas le cas, un processus à minimum de phase peut se déduire en remplaçant les zéros $\{z_{l,i}\}_{l=1,\dots,m_i \leq q_i}$ de module supérieur à 1 par $1/z_{l,i}^*$ pour $l = 1, \dots, m_i$. La fonction de transfert résultante est alors notée $H_{min,i}(z)$. Si $K_{l,i} = |z_{l,i}|^2$ pour les zéros modifiés, la variance $\sigma_{u,min,i}^2$ du processus générateur associé vaut :

$$\sigma_{u,min,i}^2 = \sigma_{u,i}^2 \prod_{l=1}^{m_i} K_{l,i} \quad (5)$$

Comme la densité de probabilité d'un vecteur X_k stockant k échantillons consécutifs d'un processus ARMA Gaussien n'est caractérisée que par son vecteur moyenne $\mu_{k,i}$ et sa matrice de covariance $Q_{k,i}$, on peut toujours considérer le processus ARMA défini par $H_{min,i}(z)$. Ce dernier peut être approximé par un modèle AR d'ordre fini $k > \max(p_i, q_i)$ comme suit :

$$x_{t,i} \approx - \sum_{j=1}^k \alpha_{j,k,i} x_{t-j,i} + u_{t,k,i} \quad (6)$$

où $u_{t,k,i}$ est le processus générateur du processus AR d'ordre k de variance $\sigma_{u,k,i}^2$ et les paramètres AR $\{\alpha_{j,k,i}\}_{j=1,\dots,k}$ associés vérifient les équations de Yule-Walker :

$$Q_{k+1,i} \begin{bmatrix} 1 \\ \alpha_{1,k,i} \\ \vdots \\ \alpha_{k,k,i} \end{bmatrix} = \begin{bmatrix} Q_{k,i} & R_{k,i} \\ R_{k,i}^T & r_{0,i} \end{bmatrix} \begin{bmatrix} 1 \\ \alpha_{1,k,i} \\ \vdots \\ \alpha_{k,k,i} \end{bmatrix} = \begin{bmatrix} \sigma_{u,k,i}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (7)$$

avec $r_{0,i}$ la fonction de covariance de $x_{t,i}$ en 0 et $R_{k,i}$ le vecteur de covariance stockant les valeurs de la fonction de covariance pour des écarts allant de k à 1.

Dans la suite, nous tiendrons compte des propriétés suivantes :
1/ $Q_{k,i}$ est non singulière même si la DSP est ponctuellement égale à 0. Seule la matrice de covariance de taille infinie ne sera pas inversible si un zéro de la fonction de transfert est de module unité.
2/ Partant de la formule du complément de Schur, $|Q_{k+1,i}|$ vaut :

$$|Q_{k+1,i}| = \begin{vmatrix} Q_{k,i} & R_{k,i} \\ R_{k,i}^T & r_{0,i} \end{vmatrix} = r_{0,i} \times |Q_{k,i} - R_{k,i} \frac{1}{r_{0,i}} R_{k,i}^T| \quad (8)$$

En utilisant des résultats sur les déterminants, on montre que :

$$\frac{|Q_{k+1,i}|}{|Q_{k,i}|} = r_{0,i} - R_{k,i}^T Q_{k,i}^{-1} R_{k,i} \stackrel{(7)}{=} \sigma_{u,k,i}^2 \quad (9)$$

Par conséquent, en faisant tendre k vers l'infini, on aboutit à :

$$\lim_{k \rightarrow +\infty} \frac{|Q_{k+1,i}|}{|Q_{k,i}|} = \sigma_{u,min,i}^2 \quad (10)$$

3/ $\sigma_{u,min,i}^2 \leq \sigma_{u,k+1,i}^2 \leq \sigma_{u,k,i}^2 \leq \sigma_{u,0,i}^2 = r_{0,i}$ pour tout $k \geq 1$ d'après Levinson [8]. Comme $Q_{1,i} = r_{0,i}$, il vient d'après (9) :

$$\sigma_{u,min,i}^{2k} < |Q_{k,i}| \leq r_{0,i}^k \quad (11)$$

4/ Soient I_k la matrice identité de taille $k \times k$, $P_{k,1} D_{k,1} P_{k,1}^T$ et $P_k D_k P_k^T$ les décompositions en valeurs propres de $(\beta - 1)Q_{k,1}$ avec $\beta \geq 1$ et de la matrice définie positive $D_{k,1}^{-\frac{1}{2}} P_{k,1}^T Q_{k,2} P_{k,1} D_{k,1}^{-\frac{1}{2}}$ et $P_{equ} = P_{k,1} D_{k,1}^{\frac{1}{2}} P_k$; le déterminant de $Q_{k,\beta-1,1}$ vérifie :

$$\begin{aligned} |Q_{k,\beta-1,1}| &= |(\beta - 1)Q_{k,1} + Q_{k,2}| \quad (12) \\ &= |P_{equ} P_{equ}^T + P_{equ} D_k P_{equ}^T| = |P_{equ}| |I_k + D_k| |P_{equ}^T| \\ &\geq |P_{equ}| (|I_k| + |D_k|) |P_{equ}^T| \\ &= (\beta - 1)^k |Q_{k,1}| + |Q_{k,2}| \underset{(11)}{>} (\beta - 1)^k \sigma_{u,min,1}^{2k} + \sigma_{u,min,2}^{2k} \end{aligned}$$

5/ On peut montrer [4] que :

$$\lim_{k \rightarrow +\infty} \left(\text{Tr}(Q_{k+1,2}^{-1} Q_{k+1,1}) - \text{Tr}(Q_{k,2}^{-1} Q_{k,1}) \right) = P^{(1,2)} \quad (13)$$

avec Tr la trace, $P^{(1,2)}$ la puissance du 1^{er} processus filtré par le filtre inverse de fonction de transfert $\frac{1}{\sigma_{u,min,2}} H_{min,2}^{-1}(z)$. A noter que si le 2^{nd} processus a un zéro de module unité qui n'est pas égal à un zéro du 1^{er} processus, $P^{(1,2)}$ est infinie.

6/ La somme pondérée de deux processus ARMA indépendants et stationnaires au sens large, $x_{t,1}$ et $x_{t,2}$, dont les ordres respectifs sont (p_1, q_1) et (p_2, q_2) , est lui aussi un processus ARMA, d'ordre (p, q) tel que $p \leq p_1 + p_2$, $q \leq \max(p_1 + q_2, p_2 + q_1)$.

2.2 BBIF Gaussien

Un processus ARMA intégré, ARIMA, d'ordre (p_i, d_i, q_i) où d_i est un entier, est une généralisation des processus ARMA d'ordre (p_i, q_i) où d_i pôles unités sont ajoutés. Ces processus sont non-stationnaires. Les processus ARFIMA d'ordre (p_i, d_i, q_i) constituent un autre niveau de généralisation où l'ordre de différenciation d_i peut prendre des valeurs réelles. Un cas particulier est le bruit blanc fractionnaire où $p_i = q_i = 0$. Sa transformée en z , $X_i(z)$ satisfait :

$$X_i(z)(1 - z^{-1})^{d_i} = U_i(z) \quad (14)$$

où $U_i(z)$ est la transformée en z du processus générateur blanc Gaussien centré et de variance $\sigma_{u,i}^2$.

1/ Etant donné (2) avec $g = -d_i$ et (14), le BBIF peut s'exprimer comme un processus à moyenne ajusté (MA) d'ordre infini dont les paramètres MA valent $\frac{\Gamma(k+d_i)}{\Gamma(k+1)\Gamma(d_i)}$ pour $k \geq 0$ et sont proportionnels pour k grand à k^{d_i-1} en s'appuyant sur la formule de Stirling. Partant de ce modèle, on montre que la variance du BBIF est finie si $d_i < \frac{1}{2}$.

2/ Etant donné (14) et si θ désigne la pulsation normalisée, la DSP du BBIF, notée $S_i(\theta)$, vérifie :

$$S_i(\theta) = \sigma_{u,i}^2 \left| 2 \sin\left(\frac{\theta}{2}\right) \right|^{-2d_i} \quad (15)$$

3/ En prenant la transformée de Fourier inverse de (15), la fonction d'autocorrélation vaut pour $-\frac{1}{2} < d_i < \frac{1}{2}$:

$$r_{\tau,i} = \frac{\Gamma(\tau + d_i)\Gamma(1 - 2d_i)}{\Gamma(\tau + 1 - d_i)\Gamma(1 - d_i)\Gamma(d_i)} \sigma_{u,i}^2 \quad (16)$$

Dès lors, $\frac{r_{\tau,i}}{r_{0,i}} = \frac{\Gamma(1-d_i)}{\Gamma(d_i)} \frac{\Gamma(\tau+d_i)}{\Gamma(\tau+1-d_i)} \approx \frac{\Gamma(1-d_i)}{\Gamma(d_i)} \tau^{2d_i-1}$ pour τ grand. Cette décroissance hyperbolique pour $0 < d_i < \frac{1}{2}$ caractérise un processus à mémoire longue. Pour $-\frac{1}{2} < d_i < 0$, le processus est dit "antipersistant" avec $r_{\tau,i} < 0$ pour $\tau > 0$.

4/ Le déterminant de la matrice de covariance $Q_{k,i}$ vérifie :

$$|Q_{k,i}| = r_{0,i}^k \prod_{\tau=1}^{k-1} (1 - \rho_{\tau,i}^2)^{k-\tau} \quad (17)$$

avec les coefficients de corrélation partielle $\rho_{\tau,i} = \frac{d_i}{\tau-d_i} \neq 1$. Comme $|\rho_{\tau,i}| < 1$, $Q_{k,i}$ est inversible. De plus, on a :

$$\lim_{k \rightarrow +\infty} \frac{|Q_{k+1,i}|}{|Q_{k,i}|} = \lim_{k \rightarrow +\infty} r_{0,i} \prod_{\tau=1}^k (1 - \rho_{\tau,i}^2) = \sigma_{u,i}^2 \quad (18)$$

En effet, le BBIF peut aussi être vu comme un processus autorégressif causal d'ordre infini quand $d_i > -\frac{1}{2}$. Le carré des coefficients de corrélation partielle coïncident alors avec le carré des coefficients de réflexion du processus, ce qui explique (18).

5/ La somme pondérée de deux BBIFs est à mémoire longue mais n'est pas un BBIF.

3 Mesures à l'étude : deux exemples

3.1 Entropie croisée de deux BBIF Gaussiens

Soit $\mathbf{p}_1(X_k)$ la densité de probabilité du vecteur colonne X_k concaténant k échantillons consécutifs du 1^{er} processus. L'entropie de Shannon, notée $H_{S,k}^{(1)}$, est définie comme suit :

$$H_{S,k}^{(1)} = - \int_{X_k} \mathbf{p}_1(X_k) \ln \mathbf{p}_1(X_k) dX_k \quad (19)$$

De plus, lorsque l'on compare deux processus aléatoires caractérisés par leur densité de probabilité respective $\mathbf{p}_1(X_k)$ et $\mathbf{p}_2(X_k)$, la divergence de KL vérifie :

$$KL_k^{(1,2)} = \int_{X_k} \mathbf{p}_1(X_k) \ln \left(\frac{\mathbf{p}_1(X_k)}{\mathbf{p}_2(X_k)} \right) dX_k \quad (20)$$

L'entropie croisée de Shannon, $H_{S,k}^{(1,2)}$, se déduit comme suit :

$$H_{S,k}^{(1,2)} = H_{S,k}^{(1)} + KL_k^{(1,2)} \quad (21)$$

Regardons leurs expressions dans le cas de vecteurs aléatoires réels Gaussiens de taille k de moyenne nulle et de matrice de covariance $Q_{k,i}$.

En tirant avantage de (19) et de (20) ainsi que de l'expression de la densité de probabilité Gaussienne, les expressions de l'entropie et de la divergence de KL deviennent :

$$\left\{ \begin{array}{l} H_{S,k}^{(1)} = \frac{k}{2} (1 + \ln(2\pi)) + \frac{1}{2} \ln |Q_{k,1}| \end{array} \right. \quad (22a)$$

$$\left\{ \begin{array}{l} KL_k^{(1,2)} = \frac{1}{2} \left(\text{Tr}(Q_{k,2}^{-1} Q_{k,1}) - k - \ln \frac{|Q_{k,1}|}{|Q_{k,2}|} \right) \end{array} \right. \quad (22b)$$

En combinant (21), (22a) et (22b), l'entropie croisée de Shannon satisfait dans le cas Gaussien :

$$H_{S,k}^{(1,2)} = \frac{k \ln(2\pi) + \text{Tr}(Q_{k,2}^{-1} Q_{k,1}) + \ln |Q_{k,2}|}{2} \quad (23)$$

Dans la suite, exprimons l'incrément, qui correspond à la différence entre cette grandeur calculée pour $k+1$ et k variables et regardons son évolution quand k augmente et tend vers l'infini. Dans ce cas, le taux d'entropie croisée que l'on note $\Delta H_S^{(1,2)} = \lim_{k \rightarrow +\infty} (H_{S,k+1}^{(1,2)} - H_{S,k}^{(1,2)})$ vérifie :

$$\Delta H_S^{(1,2)} = \frac{1}{2} \left(\ln(2\pi) + \ln \sigma_{u,2}^2 + P^{(1,2)} \right) \quad (24)$$

où $P^{(1,2)}$ désigne la puissance du 1^{er} processus filtré par le filtre inverse dont la fonction de transfert est $\frac{1}{\sigma_{u,2}} H_2^{-1}(z)$. Ce signal filtré est un BBIF d'ordre $d_1 - d_2$. La variance du processus générateur vaut $\frac{\sigma_{u,1}^2}{\sigma_{u,2}^2}$. Quand $|d_1 - d_2| < \frac{1}{2}$, la puissance

$P^{(1,2)} = \frac{\Gamma(1-2(d_1-d_2))}{\Gamma^2(1-(d_1-d_2))} \frac{\sigma_{u,1}^2}{\sigma_{u,2}^2}$ est finie. Le taux d'entropie croisée

est donc fini et étant donné l'expression de $\Delta H_S^{(1,2)}$, l'entropie croisée de Shannon peut s'approximer par une fonction affine dont la pente est $\Delta H_S^{(1,2)}$. Puisque l'entropie croisée est toujours positive ou nul, le taux d'entropie croisée est positif ou nul. Si $|d_1 - d_2| > \frac{1}{2}$, le taux est infini. L'ensemble des paramètres définissant les processus concourt à la définition du taux d'entropie croisée. A noter que les taux d'entropie de Shannon et de divergence de KL valent :

$$\left\{ \begin{array}{l} \Delta H_S^{(1)} = \frac{1}{2} (1 + \ln(2\pi)) + \frac{1}{2} \ln \sigma_{u,1}^2 \end{array} \right. \quad (25a)$$

$$\left\{ \begin{array}{l} \Delta KL_k^{(1,2)} = \frac{1}{2} \left(P^{(1,2)} - 1 - \ln \frac{\sigma_{u,1}^2}{\sigma_{u,2}^2} \right) \end{array} \right. \quad (25b)$$

Une illustration est donnée à la figure 1 où l'on observe que les taux de divergence et d'entropie croisée diminuent au fur et à mesure que d_2 se rapproche de d_1 .

3.2 β -divergence de processus ARMA centré

Parmi les familles de divergences qui généralisent la divergence de KL, regardons la β -divergence :

$$BeD_k^{(1,2)}(\beta) = \frac{1}{\beta(\beta-1)} I_2(\mathbf{p}_1, \beta) + \frac{1}{\beta} I_2(\mathbf{p}_2, \beta) - \frac{1}{\beta-1} I_1(\mathbf{p}_1, \mathbf{p}_2, 1, \beta-1) \quad (26)$$

Cette dernière se ramène à la divergence d'Itakura-Saito quand β tend vers 0. Si β tend vers 1, on retrouve la divergence de KL. Il suffit pour cela d'utiliser la règle de L'Hôpital. Quand $\beta = 2$, la divergence est symétrique et peut s'interpréter, à un facteur

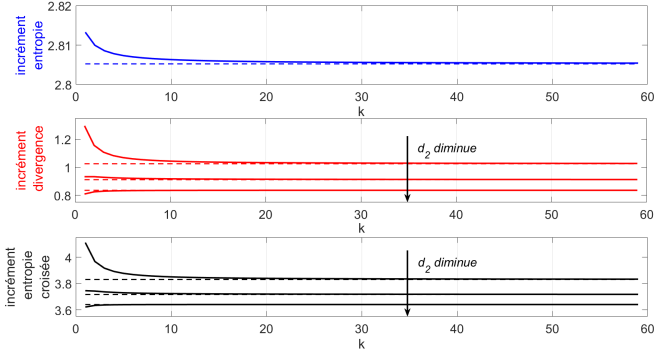


FIGURE 1 – Convergence des incréments d’entropie, divergence et entropie croisée vers leurs taux respectifs (pointillés), avec $d_1 = 0.15$, $d_2 = 0.45, 0.35$ puis 0.25 , $\sigma_{u,1} = 4$ et $\sigma_{u,2} = 2$.

multiplicatif près, comme la norme 2 de la différence entre les densités de probabilité :

$$BeD_k^{(1,2)}(2) = \frac{1}{2} \int_{X_k} (\mathbf{p}_1(X_k) - \mathbf{p}_2(X_k))^2 dX_k \quad (27)$$

Procédons comme dans le premier exemple. Etant donné (26), l’incrément de la β -divergence s’écrit comme suit :

$$\Delta BeD_k^{(1,2)}(\beta) = \frac{\delta BeD_{k,1}^{(1,2)}}{\beta(\beta-1)} + \frac{\delta BeD_{k,2}^{(1,2)}}{\beta} - \frac{\delta BeD_{k,3}^{(1,2)}}{\beta-1} \quad (28)$$

où les termes $\{\delta BeD_{k,i}^{(1,2)}\}_{i=1,2}$ vérifient :

$$\delta BeD_{k,i}^{(1,2)} = B_{k,i} \left(\frac{1}{\beta^{\frac{k}{2}} (2\pi)^{\frac{(\beta-1)}{2}}} \frac{|Q_{k,i}|^{\frac{(\beta-1)}{2}}}{|Q_{k+1,i}|^{\frac{(\beta-1)}{2}}} - 1 \right) \quad (29)$$

avec $B_{k,i} = \frac{1}{\beta^{\frac{k}{2}} (2\pi)^{\frac{(\beta-1)}{2}} |Q_{k,i}|^{\frac{(\beta-1)}{2}}}$ et le dernier terme vaut :

$$\delta BeD_{k,3}^{(1,2)} = B_{k,3} \left(\frac{|Q_{k,\beta-1,1}|^{\frac{1}{2}} |Q_{k,2}|^{\frac{\beta-2}{2}}}{(2\pi)^{\frac{(\beta-1)}{2}} |Q_{k+1,\beta-1,1}|^{\frac{1}{2}} |Q_{k+1,2}|^{\frac{\beta-2}{2}}} - 1 \right) \quad (30)$$

avec $B_{k,3} = \frac{1}{(2\pi)^{\frac{k(\beta-1)}{2}} |Q_{k,\beta-1,1}|^{\frac{1}{2}} |Q_{k,2}|^{\frac{\beta-2}{2}}}$.

Les limites de ces grandeurs quand k tend vers l’infini sont :

$$\begin{cases} \delta BeD_i^{(1,2)} = \left(\frac{1}{\beta^{\frac{k}{2}} (2\pi)^{\frac{(\beta-1)}{2}} \sigma_{u,min,i}^{\beta-1}} - 1 \right) \lim_{k \rightarrow +\infty} B_{k,i}, i = 1, 2 \\ \delta BeD_3^{(1,2)} = \left(\frac{1}{(2\pi)^{\frac{(\beta-1)}{2}} \sigma_{u,min,\beta-1,1} \sigma_{u,2}^{\beta-2}} - 1 \right) \lim_{k \rightarrow +\infty} B_{k,3} \end{cases}$$

avec $\sigma_{u,min,\beta-1,1}$ l’écart type du processus générateur du processus ARMA à minimum de phase associé à $Q_{k,\beta-1,1}$.

Comme nous allons le voir, le taux de divergence peut souvent être nul. Sans chercher à analyser les compensations entre les trois composantes $\{\delta BeD_{k,i}^{(1,2)}\}_{i=1,\dots,3}$ et l’influence de leurs poids, intéressons-nous, quand $\sigma_{u,min,1}^2 \geq \sigma_{u,min,2}^2$ et β n’est pas ”trop proche” de 1, à $(2\pi)^{\frac{k(\beta-1)}{2}} |Q_{k,\beta-1,1}|^{\frac{1}{2}} |Q_{k,2}|^{\frac{\beta-2}{2}}$, qui est le dénominateur de $B_{k,3}$. En utilisant (11) et (12), on peut le minorer et par voie de conséquence majorer $B_{k,3}$ comme suit :

$$B_{k,3} \leq \frac{1}{((2\pi)^{\frac{(\beta-1)}{2}} (\beta-1)^{\frac{1}{2}} \sigma_{u,min,2}^{\beta-1})^k} = T_k \quad (31)$$

Comme $(2\pi)^{\frac{(\beta-1)}{2}} \beta^{\frac{1}{2}} \sigma_{u,min,i}^{\beta-1} \geq (2\pi)^{\frac{(\beta-1)}{2}} (\beta-1)^{\frac{1}{2}} \sigma_{u,min,2}^{\beta-1}$, on a :

$$B_{k,i} \leq \frac{1}{((2\pi)^{\frac{(\beta-1)}{2}} \beta^{\frac{1}{2}} \sigma_{u,min,2}^{\beta-1})^k} \leq T_k \text{ avec } i = 1, 2 \quad (32)$$

Ainsi, si $\sigma_{u,min,1}^2 \geq \sigma_{u,min,2}^2 > \frac{1}{2\pi(\beta-1)^{\frac{1}{2}}}$, $\lim_{k \rightarrow +\infty} B_{k,i} = 0$ pour $i = 1, 3$. Le taux de divergence est donc nul. Cf. figure 2.

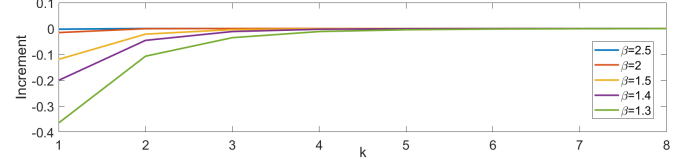


FIGURE 2 – Convergence des incréments de β -divergence, pour deux processus ARMA d’ordre (4,6). Pour le 1^{er} , les zéros et les pôles sont respectivement : $0.8e^{j\frac{\pi}{4}}$ et $0.9e^{j\frac{3\pi}{4}}$ et leurs conjugués. $0.9e^{j\frac{\pi}{6}}$, $0.7e^{j\frac{\pi}{5}}$ et $0.9e^{j\frac{\pi}{4}}$ et leurs conjugués. $\sigma_{u,1}^2 = 25$. Pour le 2^{nd} , les zéros et les pôles sont : $0.8e^{j\frac{\pi}{3}}$ et $0.6e^{j\frac{3\pi}{5}}$ et leurs conjugués. $0.95e^{j\frac{5\pi}{6}}$, $0.7e^{j\frac{3\pi}{5}}$ et $0.9e^{j\frac{\pi}{4}}$ et leurs conjugués. $\sigma_{u,2}^2 = 16$.

4 Conclusions et perspectives

Le taux d’entropie croisée entre BBIF Gaussiens peut être adapté pour mettre en avant leur dissimilarité. Quant au taux de β -divergence entre processus ARMA Gaussiens centrés, il s’avère souvent nul et ne semble donc pas adapté à la comparaison de ces processus. A l’avenir, nous comptons définir le taux d’une autre manière pour ce type de divergence.

Références

- [1] G. Ciuperca *et al.* Computation and estimation of generalized entropy rates for denumerable Markov chains. IEEE Trans. on Information Theory, 57 :4026–4034, 2011.
- [2] M. Gil. On Rényi divergence measures for continuous alphabet sources. PhD Thesis, Queen’s University, 2011.
- [3] V. Girardin *et al.* Different closed-form expressions for generalized entropy rates of Markov chains. Methodology and Computing in Applied Probability, 21 :1431–1452, 2019.
- [4] E. Grivel *et al.* Interpreting the asymptotic increment of Jeffrey’s divergence between some random processes. Digital Signal Processing, 75, (4) :120–133, 2018.
- [5] E. Grivel *et al.* Kullback-Leibler and Rényi divergence rate for Gaussian stationary ARMA processes comparison. Digital Signal Processing, 116 :103089, 2021.
- [6] E. Grivel. Studying three families of divergences to compare wide-sense stationary Gaussian ARMA processes. ICASSP 2022.
- [7] P. Kluzka. On Jensen-Rényi and Jeffreys-Rényi type f-divergences induced by convex functions. Physica A : Statistical Mechanics and its Applications, 548 :122527, 2020.
- [8] M. Najim, Modélisation, estimation et filtrage optimal en traitement du signal, Lavoisier.
- [9] F. Nielsen *et al.* A closed-form expression for the Sharma-Mittal entropy of exponential families. Journal of Physics A : Mathematical and Theoretical, 45, 2012.
- [10] Z. Rached *et al.* The Kullback–Leibler divergence rate between Markov sources. IEEE Trans. on Information Theory, 50 :917–921, 2004.
- [11] P. Regnault, *et al.* Weighted closed form expressions based on escort distributions for Rényi entropy rates of Markov chains. GSI’17, 2017.
- [12] R. F. Vigelis, *et al.* Conditions for the existence of a generalization of Rényi divergence. Physica A, 558 :124953, 2020.