

# Contribution à l'estimation de l'entropie de transfert par les plus proches voisins

JIE ZHU<sup>1,2,3</sup>, JEAN-JACQUES BELLANGER<sup>1,2</sup>, REGINE LE BOUQUIN JEANNES<sup>1,2,3</sup>

<sup>1</sup> INSERM, U 1099, Rennes, F-35000, France

<sup>2</sup> Université de Rennes 1, LTSI, F-35000, France

<sup>3</sup> Centre de Recherche en Information Biomédicale sino-français (CRIBs), Rennes, France  
Université de Rennes 1, Campus de Beaulieu, 35042 Rennes cedex, FRANCE

<sup>1</sup>regine.le-bouquin-jeannes@univ-rennes1.fr

**Résumé** - Ce papier présente une contribution à l'estimation de l'entropie d'un vecteur aléatoire par la méthode des  $k$  plus proches voisins ( $k$  PPV) et à son utilisation pour mesurer l'entropie de transfert, la question centrale étant celle du biais. Pour estimer l'entropie nous proposons d'utiliser des voisinages (hyper-)rectangulaires de volume minimal incluant les  $k$  PPV à la place de voisinages (hyper-)cubiques. Cette approche est mise en œuvre en étendant l'estimateur de Kozachenko-Leonenko pour aboutir à une nouvelle mesure d'entropie de transfert. Des résultats expérimentaux permettent une comparaison avec les deux estimateurs d'entropie de transfert basés sur les  $k$  PPV les plus usités dans la littérature, dérivés de l'estimateur d'information mutuelle de Kraskov-Stögbauer-Grassberger.

**Abstract** - This paper contributes to the estimation of entropy of a random vector using the  $k$ -nearest neighbors ( $k$  NN) based technique and to its use to measure transfer entropy, focalizing on bias reduction. To estimate entropy, we propose to use (hyper-)rectangular neighborhoods with minimal volume including these  $k$  NN instead of standard (hyper-)cubic neighborhoods. To this end, we extended the standard Kozachenko-Leonenko entropy estimator to get a new transfer entropy measure. Experimental simulations allow us to compare the new estimator with two estimators available in free toolboxes, both based on the Kraskov-Stögbauer-Grassberger mutual information estimator.

## 1 Introduction

L'entropie de transfert est une mesure statistique non paramétrique, définie à partir du concept d'entropie de Shannon, qui vise à mesurer un débit d'information dirigée depuis un premier système dynamique  $\mathcal{A}$  vers un deuxième système  $\mathcal{B}$ . Etant donnée l'évolution passée de  $\mathcal{A}$ , l'entropie de transfert de  $\mathcal{B}$  vers  $\mathcal{A}$  est définie comme la réduction d'incertitude (au sens de Shannon) sur l'évolution future de  $\mathcal{A}$  lorsque l'on prend en compte la connaissance du passé de  $\mathcal{B}$  en complément de celle du passé de  $\mathcal{A}$ . Introduite par Schreiber [1], cette grandeur a connu et/ou connaît des applications dans différents domaines, comme la physique, la physiologie, en particulier les neurosciences [2]. D'un point de vue pratique on ne connaît l'évolution des états de  $\mathcal{A}$  et  $\mathcal{B}$  que par l'intermédiaire de capteurs délivrant des signaux d'observation, disons, après échantillonnage,  $X_i \in \mathbb{R}$ ,  $i \in \mathbb{Z}$  pour  $\mathcal{A}$  et  $Y_i \in \mathbb{R}$ ,  $i \in \mathbb{Z}$  pour  $\mathcal{B}$ . Inférer sur l'éventuelle influence causale de  $\mathcal{A}$  sur  $\mathcal{B}$  passe donc par l'élaboration d'une statistique sur cette paire de signaux, l'entropie de transfert, proposée dans [1], étant celle que nous considérons. Définissons  $U_i^{(k)} = (U_{i-k+1}, U_{i-k+2}, \dots, U_i)$  où  $U$  peut correspondre à  $X$  ou à  $Y$ . Pour une paire donnée  $(m, n)$  d'entiers positifs, on définit l'entropie de transfert, TE (Transfer Entropy), comme différence de deux entropies conditionnelles :

$$\text{TE}_{Y \rightarrow X, i} = \mathcal{H}(X_i^p | X_i^-) - \mathcal{H}(X_i^p | X_i^-, Y_i^-) \quad (1)$$

où  $X_i^p = X_{i+1}$ ,  $X_i^- \triangleq X_i^{(m)}$ ,  $Y_i^- \triangleq Y_i^{(n)}$  sont des vecteurs aléatoires de dimensions respectives 1,  $m$ ,  $n$ . La mesure de probabilité jointe  $\mathcal{P}_{X_i^p, X_i^-, Y_i^-}$  est supposée ici continue (par rapport à la mesure de Lebesgue idoine), de même que les mesures conditionnelles  $\mathcal{P}_{X_i^p | X_i^-, Y_i^-}$ ,  $\mathcal{P}_{X_i^p | X_i^-}$ . Si  $Y$  n'a aucune influence sur  $X$ , alors  $\mathcal{P}_{X_i^p | X_i^-, Y_i^-} \equiv \mathcal{P}_{X_i^p | X_i^-}$  et  $\text{TE}_{Y \rightarrow X, i} = 0$ . Quand  $X$  et  $Y$  sont conjointement stationnaires, les vecteurs  $(X_i^p, X_i^-, Y_i^-)$  ont même loi qu'un vecteur noté  $(X^p, X^-, Y^-)$ . Bien que primordial dans les applications, le choix de  $(m, n)$  n'est pas considéré ici. On notera  $\mathcal{S}_{X^p, X^-, Y^-}$ ,  $\mathcal{S}_{X^p, X^-}$ ,  $\mathcal{S}_{X^-, Y^-}$  et  $\mathcal{S}_{X^-}$  les espaces dans lesquels  $(X^p, X^-, Y^-)$ ,  $(X^p, X^-)$ ,  $(X^-, Y^-)$ , et  $X^-$  prennent respectivement leurs valeurs. L'observation disponible sera d'une longueur de  $N$  échantillons,  $(X_i, Y_i)$ ,  $i = 1, \dots, N$ . Les symboles  $p_U$  et  $\mathcal{H}(U)$  désignant respectivement la densité de probabilité et l'entropie différentielle de Shannon (en nats, "log" étant le logarithme népérien) pour un vecteur  $U$ , l'entropie de transfert peut également se réécrire sous la forme de quatre entropies différentielles,

$$\begin{aligned}
& \text{TE}_{Y \rightarrow X, i} \\
&= -E \left[ \log \left( p_{X_i^-, Y_i^-} \left( X_i^-, Y_i^- \right) \right) \right] - E \left[ \log \left( p_{X_i^P, X_i^-} \left( X_i^P, X_i^- \right) \right) \right] \\
&\quad + E \left[ \log \left( p_{X_i^P, X_i^-, Y_i^-} \left( X_i^P, X_i^-, Y_i^- \right) \right) \right] + E \left[ \log \left( p_{X_i^-} \left( X_i^- \right) \right) \right] \\
&= \mathcal{H} \left( X_i^-, Y_i^- \right) + \mathcal{H} \left( X_i^P, X_i^- \right) - \mathcal{H} \left( X_i^P, X_i^-, Y_i^- \right) - \mathcal{H} \left( X_i^- \right)
\end{aligned} \tag{2}$$

Une structure naturelle d'estimateur de TE correspond alors à une somme de quatre estimateurs d'entropie :

$$\begin{aligned}
\widehat{\text{TE}}_{Y \rightarrow X} &= \widehat{\mathcal{H} \left( X^-, Y^- \right)} + \widehat{\mathcal{H} \left( X^P, X^- \right)} \\
&\quad - \widehat{\mathcal{H} \left( X^P, X^-, Y^- \right)} - \widehat{\mathcal{H} \left( X^P \right)}
\end{aligned} \tag{3}$$

Dans la section suivante nous présentons les estimateurs d'entropie et de transfert d'entropie qui nous servent de référence et les nouveaux estimateurs correspondants.

## 2 Estimations d'entropie et de TE

### 2.1 Etat de l'art

Soit  $x_i, i=1, \dots, N$  in  $\mathbb{R}^{d_x}$  une réalisation d'une suite iid  $X_1, \dots, X_N$ . L'ensemble des  $k$  plus proches voisins (PPV) de  $x_i$  dans cette réalisation et la distance entre  $x_i$  et son  $k^{\text{ème}}$  PPV sont notés respectivement  $\mathcal{X}_i^k$  et  $d_{x_i, k}$ . Soit  $\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right) \subset \mathbb{R}^{d_x}$  un voisinage de  $x_i$  dans  $\mathbb{R}^{d_x}$  incluant  $x_i$  et  $\mathcal{X}_i^k$ . Pour une norme  $\|\cdot\|$  sur  $\mathbb{R}^{d_x}$  (Euclidienne, du sup, ...) une construction standard de  $\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right)$  est l'(hyper-)balle de rayon  $d_{x_i, k}$ , i.e.  $\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right) = \{x : \|x - x_i\| \leq d_{x_i, k}\}$  d'(hyper-)volume, i.e. de mesure de Lebesgue,  $v_i = \int_{\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right)} dx$ .

L'estimateur d'entropie de Kozachenko-Leonenko est défini par (voir par exemple (10) dans [3])

$$\widehat{\mathcal{H} \left( X \right)}_{KL} = \psi(N) + \frac{1}{N} \sum_{i=1}^N \log(v_i) - \psi(k) \tag{4}$$

où  $v_i$  est le volume de  $\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right) = \{x : \|x - x_i\| \leq d_{x_i, k}\}$  obtenu avec la norme du sup et où  $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$  est la fonction digamma. Son obtention, résumée dans [4], introduit  $h^{x_i}(\varepsilon) \triangleq \int_{\|u - x_i\| \leq \varepsilon/2} dP_X(u)$ , qui peut être approchée par  $p_X(x_i) \int_{\|u\| \leq \varepsilon/2} du = p_X(x_i) c_d \varepsilon^d$  pour  $\varepsilon \searrow 0$  (d'après un lemme de Lebesgue, [3]). On vérifie que  $E \left[ \log \left( h^{X_i} \left( D_{X_i, k} \right) \right) \right] = \psi(k) - \psi(N)$  où  $h^{X_i}$  et  $D_{X_i, k}$  sont les quantités aléatoires associées à  $h^{x_i}(\cdot)$  et  $d_{x_i, k}$ , et où  $c_d$  est le volume de la boule  $d$ -

dimensionnelle de rayon unité. On en déduit (4) en approximant l'opérateur espérance mathématique par une moyenne temporelle et en identifiant. Moyennant quelques hypothèses, l'estimation (4) s'avère être asymptotiquement sans biais. Basé sur (3) et (4) il a été proposé dans [5] un estimateur de l'entropie de transfert ici qualifié de standard et noté  $\widehat{\text{TE}}_{Y \rightarrow X, S}$

$$\begin{aligned}
\widehat{\text{TE}}_{Y \rightarrow X, S} &= \psi(k) + \frac{1}{N} \sum_{i=1}^N \left[ \psi(n_{X_i^-, i} + 1) \right. \\
&\quad \left. - \psi(n_{(X^-, Y^-), i} + 1) - \psi(n_{(X^P, X^-), i} + 1) \right]
\end{aligned} \tag{5}$$

obtenu en (i) sélectionnant  $k$  PPV avec la norme du sup dans  $\mathcal{S}_{X^P, X^-, Y^-}$ , (ii) utilisant, pour chaque  $i$ , la distance  $\varepsilon_i \triangleq d_{(x_i^P, x_i^-, y_i^-), k}$  pour construire des boules de même rayon dans les 3 autres sous-espaces de dimensions inférieures,  $\mathcal{S}_{X^-}$ ,  $\mathcal{S}_{X^-, Y^-}$  et  $\mathcal{S}_{X^P, X^-}$ , ce qui entraîne une compensation des termes correspondant au logarithme dans (4). Les trois nombres de points, parmi les  $N-1$  possibles, strictement inclus dans chacune de ces boules (en fait des (hyper-)cubes) correspondent à des entiers  $n_{X_i^-, i}$ ,  $n_{(X^-, Y^-), i}$ ,  $n_{(X^P, X^-), i}$  non égaux. L'estimateur défini en (5) a été étendu par [6, 7] en remplaçant les (hyper-)cubes dans  $\mathcal{S}_{X^-, Y^-}$  et  $\mathcal{S}_{X^P, X^-}$  par des produits de deux (hyper-)cubes de tailles différentes. Sans détailler plus avant cet algorithme, nous le dénommons par la suite algorithme étendu et l'écrivons

$$\begin{aligned}
\widehat{\text{TE}}_{Y \rightarrow X, E} &= \frac{1}{N} \sum_{i=1}^N \left( \psi(k) - \frac{2}{k} + \psi(l_{X_i^-, i}) - \psi(l_{(X^P, X^-), i}) \right. \\
&\quad \left. + \frac{1}{l_{(X^P, X^-), i}} - \psi(l_{(X^-, Y^-), i}) + \frac{1}{l_{(X^-, Y^-), i}} \right)
\end{aligned} \tag{6}$$

où les  $l_{\cdot, i}$  correspondent, par exemple pour  $l_{(X^P, X^-), i}$ , au nombre de points inclus dans un (hyper-)rectangle produit cartésien de deux (hyper-)cubes, le premier dans  $\mathcal{S}_{X^P}$  et le second dans  $\mathcal{S}_{X^-}$ . Les longueurs d'arrête  $d_{x_i^-, i}^{\max}$  et  $d_{x_i^-, i}^{\max}$  sont définies semblablement comme suit  $d_{x_i^-, i}^{\max} = 2 \times \max \left\{ \|x_k^P - x_i^P\| : (x^P, x^-, y^-)_k \in \mathcal{X}_{(x^P, x^-, y^-), i}^k \right\}$  et  $l_{(X^P, X^-), i}$  est égal au cardinal de l'ensemble :

$$\left\{ \left( x_j^P, x_i^- \right)_{1 \leq j \leq N, j \neq i} : \|x_j^P - x_i^P\| \leq d_{x_i^-, i}^{\max}, \|x_j^- - x_i^-\| \leq d_{x_i^-, i}^{\max} \right\}.$$

### 2.2 Méthode proposée

L'estimateur de l'entropie  $\mathcal{H}(X)$  proposé ici est basé sur une construction différente de  $\mathcal{D}_{x_i} \left( \mathcal{X}_i^k \right)$  qui n'est

plus l'hypercube de rayon (norme du sup)  $d_{x_i,k}$ , étant remplacé par l'(hyper-)rectangle, de centre de symétrie  $x_i$ , de volume minimal incluant  $\mathcal{X}_i^k$  (non strictement). En considérant le vecteur aléatoire  $(\epsilon_1, \dots, \epsilon_d)$  des  $d$  largeurs de cet (hyper-)rectangle dont on observe la réalisation  $(\epsilon_1, \dots, \epsilon_d) \in \mathbb{R}^d$ , les demi-largeurs sont facilement déterminées en recherchant pour chaque direction de  $\mathbb{R}^d$  à quelle distance maximale se projette l'ensemble  $\mathcal{X}_i^k$ . Il est donné dans [4] le résultat

$$E\left[\log\left(h^{x_i}(\epsilon_1, \dots, \epsilon_d)\right)\right] = \psi(k) - \frac{d-1}{k} - \psi(N) \quad (7)$$

A partir de (6), et toujours en remplaçant l'espérance par une moyenne temporelle, et par identification, nous avons obtenu le nouvel estimateur d'entropie  $\widehat{\mathcal{H}(X)}_R$

$$\widehat{\text{TE}}_{Y \rightarrow X^p} = \frac{1}{N} \sum_{i=1}^N \log \frac{v_{(X^p, X^-), i} \cdot v_{(X^-, Y^-), i}}{v_{(X^p, X^-, Y^-), i} \cdot v_{X^-, i}} + \frac{1}{N} \sum_{i=1}^N \left( \psi(k) + \psi(k_{X^-, i}) - \psi(k_{(X^p, X^-), i}) - \psi(k_{(X^-, Y^-), i}) \right. \\ \left. + \frac{d_{X^p} + d_{X^-} - 1}{k_{(X^p, X^-), i}} + \frac{d_{X^-} + d_{Y^-} - 1}{k_{(X^-, Y^-), i}} - \frac{d_{X^p} + d_{X^-} + d_{Y^-} - 1}{k} - \frac{d_{X^-} - 1}{k_{X^-, i}} \right) \quad (9)$$

### 3 Expérimentations

#### 3.1 Modèles linéaires

Le premier modèle linéaire (Modèle 1) que nous avons choisi d'expérimenter est le suivant :

$$X_t = aY_t + bZ_t + W_t, \quad W_t \in \mathbb{R}, \quad Y \in \mathbb{R}^{d_Y}, \quad Z \in \mathbb{R}^{d_Z} \quad (10)$$

où  $Y_t \sim \mathcal{N}(0, C_Y)$ ,  $Z_t \sim \mathcal{N}(0, C_Z)$ ,  $W_t \sim \mathcal{N}(0, \sigma_W^2)$ , les trois processus  $Y$ ,  $Z$ , et  $W$  étant mutuellement indépendants. Le triplet  $(X_t, Y_t, Z_t)$  correspond au triplet  $(X_t^p, X_t^-, Y_t^-)$  introduit précédemment. Même si ce modèle n'exhibe pas le passé et le futur de séries chronologiques, sa structure introduit toutes les difficultés susceptibles d'être rencontrées pour estimer une entropie de transfert, si ce n'est qu'en général  $(X_t^p, X_t^-, Y_t^-)$  n'est pas iid, comme pour le modèle AR introduit plus bas.  $C_U$  est une matrice Toeplitz dont la première ligne est égale à  $[1, \alpha, \dots, \alpha^{d_U-1}]$ . Pour la matrice  $C_Y$ , nous avons choisi  $\alpha = 0,5$ , et, pour  $C_Z$ ,  $\alpha = 0,2$ . L'écart-type  $\sigma_W$  est fixé à  $0,5$ . Les vecteurs  $a$  et  $b$  sont tels que  $a = 0,1 * [1, 2, \dots, d_Y]$  et  $b = 0,1 * [d_Z, d_Z - 1, \dots, 1]$ . A partir de ce modèle, on cherche à estimer la somme d'entropies  $\mathcal{H}(Y, Z) + \mathcal{H}(X, Y) - \mathcal{H}(X, Y, Z) - \mathcal{H}(Y)$  (qui correspond, avec les notations de l'introduction, à

$$\widehat{\mathcal{H}(X)}_R = \psi(N) + \frac{1}{N} \sum_{i=1}^N \log(v_i) - \psi(k) + \frac{d-1}{k} \quad (8)$$

(asymptotiquement non biaisé comme pour (4)) où  $v_i$  est le volume minimal de l'hyper rectangle centré sur le point  $x_i$ . Finalement, par application de (3) on en déduit le nouvel estimateur de TE,  $\widehat{\text{TE}}_{Y \rightarrow X^p}$  (9), où  $d_{X^p} = \dim(\mathcal{S}_{X^p})$ ,  $d_{X^-} = \dim(\mathcal{S}_{X^-})$ ,  $d_{Y^-} = \dim(\mathcal{S}_{Y^-})$ , où les  $v_{i,j}$  désignent les volumes des (hyper-)rectangles de volume minimal, et où les  $k_{(c),i}$  sont les nombres de points inclus non strictement dans les 4 hyper-rectangles minimaux obtenus dans  $\mathcal{S}_{X^-}$ ,  $\mathcal{S}_{X^-, Y^-}$ ,  $\mathcal{S}_{X^p, X^-}$  et  $\mathcal{S}_{X^p, X^-, Y^-}$ .

$\mathcal{H}(X_t^-, Y_t^-) + \mathcal{H}(X_t^p, X_t^-) - \mathcal{H}(X_t^p, X_t^-, Y_t^-) - \mathcal{H}(X_t^-)$  ) pour tester si la connaissance des signaux  $Y$  et  $Z$  peut améliorer la prédiction de  $X$  comparé à la seule connaissance de  $Y$ .

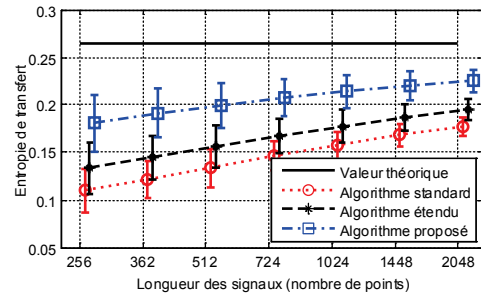


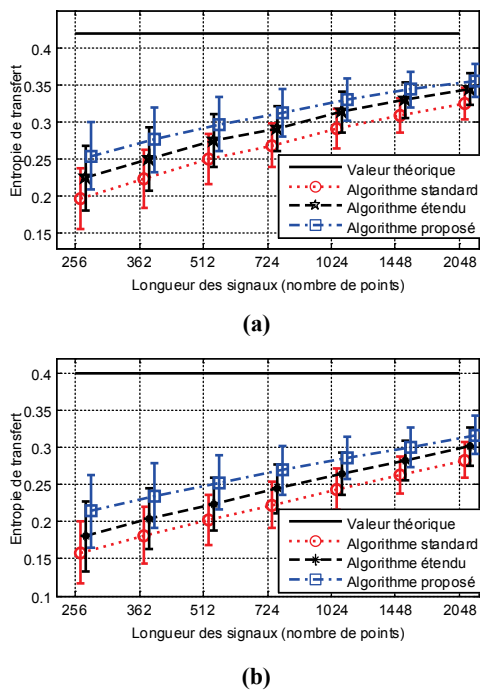
Figure 1. Entropie de transfert de  $Z$  vers  $X$  (Modèle 1) estimée pour une dimension  $d_Y = d_Z = 3$  et  $k = 8$ . La figure indique les valeurs moyennes et les écarts-types.

Si les trois algorithmes testés tendent vers la valeur théorique de TE (égale à  $0,265$ ), avec des variances comparables, l'approche proposée s'avère plus performante en termes de biais que les deux algorithmes originaux.

Dans un second temps, nous avons testé le modèle Auto-Régressif Vectoriel (VAR) suivant (Modèle 2) :

$$\begin{cases} x_t = 0,45\sqrt{2}x_{t-1} - 0,9x_{t-2} - 0,6y_{t-2} + e_{x,t} \\ y_t = 0,6x_{t-2} - 0,175\sqrt{2}y_{t-1} + 0,55\sqrt{2}y_{t-2} + e_{y,t} \end{cases} \quad (11)$$

Pour ce modèle bidirectionnel,  $e_x$  et  $e_y$  représentent les réalisations de bruits blancs gaussiens indépendants de moyenne nulle et de variance 0,1.



**Figure 2. Entropie de transfert (a) de  $X$  vers  $Y$  et (b) de  $Y$  vers  $X$  (Modèle 2) estimé pour  $k=8$ . La figure indique les valeurs moyennes et les écarts-types.**

Les résultats obtenus sont présentés sur la Figure 2. Les valeurs théoriques de  $TE_{X \rightarrow Y}$  et  $TE_{Y \rightarrow X}$  sont égales respectivement à 0,420 et 0,399. Les trois estimateurs testés tendent d'autant plus vers ces valeurs que la longueur des signaux est grande, le nouvel algorithme attestant du biais le plus faible, les variances des 3 estimateurs étant très proches.

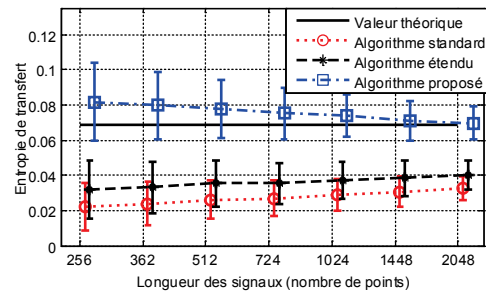
### 3.2 Modèle non linéaire

Dans un second temps, nous avons testé un modèle non linéaire (Modèle 3) obtenu en modifiant le premier modèle linéaire précédent (Modèle 1) :

$$X_t = aY_t + bS_t Z_t + W_t, \quad W_t \in \mathbb{R}, \quad Y \in \mathbb{R}^{d_Y}, \quad Z \in \mathbb{R}^{d_Z} \quad (12)$$

où  $S_t$  correspond à une suite aléatoire iid à valeurs dans  $\{-1, 1\}$ . L'intérêt de ce modèle est de pouvoir tester nos estimateurs pour  $d_Y$  et  $d_Z$  supérieurs à 1 dans un cas non gaussien tout en étant capables de calculer numériquement avec une bonne précision la valeur théorique de TE. Pour ce modèle, avec  $P(S_t = 1) = 1/2$ , on vérifie que l'indice de Granger [8] est nul bien que la mesure de TE soit strictement positive, l'indice de Granger s'avère donc effectivement inopérant ici. La Figure 3 montre que, dans ce cas non linéaire, le nouvel estimateur converge beaucoup plus rapidement et plus précisément vers la valeur théorique (égale à

0,069) que les deux algorithmes de référence, pour une variance très légèrement supérieure.



**Figure 3. Entropie de transfert de  $Z$  vers  $X$  (Modèle 3) estimé pour une dimension  $d_Y = d_Z = 3$  et  $k=8$ . La figure indique les valeurs moyennes et les écarts-types.**

## 4 Conclusion

Nous avons proposé un nouvel estimateur de l'entropie de transfert basé sur la méthode des plus proches voisins en introduisant un voisinage rectangulaire. Testé sur des modèles linéaires et non linéaires, pour des temps de calculs comparables, celui-ci s'avère plus précis que les méthodes issues de la littérature.

## 5 Références

- Schreiber, T. Measuring information transfer. *Physical review letters*, 2000, 85, 461.
- Yang, C.; Le Bouquin Jeannès, R.; Bellanger, J. J.; Shu, H. A new strategy for model order identification and its application to transfer entropy for EEG signals analysis. *Biomedical Engineering, IEEE Transactions on*, 2013, 60(5), 1318-1327.
- Leonenko N.; Pronzato L.; Savani V. A class of Rényi information estimators for multidimensional densities. *The Annals of Statistics*, 2008, 36(5), 2153-2182.
- Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Physical review E*, 2004, 69, 066138.
- Lindner, M.; Vicente, R.; Priesemann, V.; Wibral, M. TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC neuroscience*, 2011, 12, 119.
- Wibral, M.; Vicente, R.; Lindner, M. Transfer Entropy in Neuroscience. In *Directed Information Measures in Neuroscience*, Springer, 2014, pp. 3-36.
- Lizier, J. T. JIDT: an information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 2014, 1, 11.
- Granger, Clive WJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* (1969): 424-438.