

Un estimateur du taux d'entropie basé sur l'Information Mutuelle.

C. Granero-Belinchón, S.G. Roux, N.B. Garnier *

Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France.

prenom.nom@ens-lyon.fr, name.surname@ens-lyon.fr

Résumé – La théorie de l'information offre des outils capables de caractériser les variables aléatoires et leurs relations. On se propose ici d'étudier comment le taux d'entropie peut caractériser les processus aléatoires stationnaires. Dans le cas Gaussien, le taux d'entropie est directement relié à la fonction de corrélation du processus mais cette relation n'est plus nécessairement valide quand le processus est non-Gaussien. Dans ce cas le taux d'entropie prend en compte tous les moments de la distribution marginale du signal mais aussi toute la structure de corrélation du processus. La connaissance et l'analyse du taux d'entropie sont donc susceptibles d'apporter des informations supplémentaires sur la nature du processus. Nous présentons deux estimateurs du taux d'entropie basés sur la recherche de plus proches voisins (k-NN) et étudions leurs performances pour la caractérisation de processus non-Gaussiens.

Abstract – Information Theory provides a framework to characterize random variables and their relationships. In this paper we describe how the entropy rate allows one to characterize stationary random processes. When the process has a Gaussian distribution, the entropy rate is directly related to the correlation function. This does not hold anymore when the process has non-Gaussian statistics, in that case, entropy rate depends *a priori* on all moments of the marginal distribution, as well as on all the correlation structure. A detailed analysis of the entropy rate is therefore susceptible to bring additional knowledge of the process. We present here two estimates of the entropy rate based on nearest neighbors search (k-NN) and detail their performance in characterizing non-Gaussian processes.

1 Motivation et objectives

L'analyse de systèmes et signaux complexes révèle très souvent des phénomènes se déployant sur de multiples échelles de temps ou d'espace. Afin d'étudier l'évolution de ces phénomènes à travers les échelles, de nombreux outils (Fourrier, analyse multi fractale, etc) ont été utilisés. D'autre part, la théorie de l'information, développée par Shannon en 1948 [1], a démontré son utilité la caractérisation des variables aléatoires. Elle a été utilisée avec succès dans un grand nombre de domaines différents telle que l'économie [2], la médecine[3], le codage, l'analyse d'image [4], ...

Le taux d'entropie, directement basé sur l'entropie de Shannon, quantifie l'information apportée par un nouveau point qui n'est pas contenue dans l'information de m points précédemment connus. Si les $m + 1$ points sont obtenus par échantillonnage uniforme du processus, alors le taux d'entropie dépend directement de l'échelle d'échantillonnage. Ainsi, en variant cette échelle on peut espérer sonder l'évolution de la structure de corrélation à travers les échelles. Cette évolution est plus riche que celle obtenue pour la fonction de corrélation dans le sens que cette dernière quantité ne prend en compte que le moment d'ordre 2 de la marginale du processus alors que le taux d'entropie prend en compte tous les moments. Il faut noter que des outils tels que Approximate Entropy [5] ou Sample Entropy [6] ne sont rien d'autre que des estimateurs du taux d'entropie de Shannon ou de Renyi d'ordre 2. En général ces

estimations se font avec une échelle d'échantillonnage fixée souvent prise égale à l'échelle d'acquisition. Afin de sonder l'évolution à travers les échelles, une quantité, appelée entropie multi-échelles (Multiscale Entropy, MSE) [7] a été proposée qui travaille non plus directement sur le signal mais sur différentes versions agrégées du signal.

Nous proposons ici de décrire l'évolution du taux d'entropie à travers les échelles, non pas à partir de différentes versions agrégées du signal mais simplement en faisant varier la distance entre deux points consécutifs. Cette quantité est sans doute un meilleur candidat pour la caractérisation de processus non Gaussiens qui possèdent des statistiques d'ordre deux identiques: comme nous allons le montrer le taux d'entropie peut être directement relié à la structure de corrélation du signal. De plus, nous proposons deux estimateurs du taux d'entropie basés sur la recherche de plus proches voisins (k-NN). Nous comparons ensuite leurs performances sur des variables corrélées de statistique Gaussienne. Nous illustrons la supériorité de l'un d'eux pour estimer l'évolution du taux d'entropie à travers les échelles.

2 Entropie et taux d'entropie

La théorie de l'information repose sur le concept d'entropie, redéfini par Shannon [1], à partir de laquelle est construit le taux entropie, h , qui mesure l'information apporté par le $m + 1$ -ème point d'un signal quand les m précédents sont connus. Sa définition habituelle repose sur une procédure de plongement temporel [8] qui construit un signal de dimension m , $X_t^{(m,\tau)}$,

*Ce travail a été réalisé grâce au soutien financier du LABEX iMUST (ANR-10-LABX-0064) de l'Université de Lyon, dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) géré par l'Agence Nationale de la Recherche (ANR).

à partir d'une signal temporel X :

$$\mathbf{x}_t^{(m,\tau)} = (x(t), x(t-\tau), x(t-2\tau), \dots, x(t-(m-1)\tau)).$$

Cette définition dépend évidemment de l'échelle de sous échantillonnage τ qui correspond à l'échelle d'observation. Augmenter la dimension de plongement m augmente a priori l'ordre des statistiques que l'on prend en compte.

2.1 Définitions

L'entropie de Shannon d'une signal réel plongé dans un espace de dimension m , $X_t^{(m,\tau)}$, se définit par une fonctionnelle de la densité de probabilité $p(\mathbf{x}_t^{(m,\tau)})$:

$$\begin{aligned} H^{(m,\tau)}(X) &= H(\mathbf{x}_t^{(m,\tau)}) \\ &= - \int_{\mathbb{R}^m} p(\mathbf{x}_t^{(m,\tau)}) \log p(\mathbf{x}_t^{(m,\tau)}) d\mathbf{x}_t^{(m,\tau)}. \end{aligned}$$

Par sa définition, l'entropie de Shannon dépend de tous les moments sauf du moment d'ordre 1 de la pdf $p(\mathbf{x}_t^{(m,\tau)})$. Elle dépend aussi de la structure de corrélations à m points. Dans le cas $m = 1$ il n'y a pas de dépendance de l'entropie en τ et on écrira $H(X) \equiv H^{(1,\tau)}(X)$.

Le taux entropie d'ordre m est un estimateur du taux d'entropie qui mesure la variation d'entropie de Shannon entre $X_t^{(m+1,\tau)}$ et $X_t^{(m,\tau)}$, deux plongements successifs du processus stochastique X :

$$h^{(m,\tau)}(X) = H^{(m+1,\tau)}(X) - H^{(m,\tau)}(X). \quad (1)$$

Comme l'entropie, le taux entropie dépend des statistiques de tout les ordres du signal X , mais eq.(1) suggère que pour une dimension $m > 1$, elle se concentre sur l'information nouvelle contenu dans les statistiques d'ordre $m + 1$.

L'information mutuelle mesure l'information partagée par deux processus $X_t^{(m,\tau)}$ et $Y_{t'}^{(p,\tau)}$. Elle est définie de la manière suivante :

$$\begin{aligned} I^{(m,p,\tau)}(Y, X)(t' - t) &= I(\mathbf{y}_{t'}^{(p,\tau)}, \mathbf{x}_t^{(m,\tau)}) \\ &= \int_{\mathbb{R}^{m+p}} p(\mathbf{y}_{t'}^{(p,\tau)}, \mathbf{x}_t^{(m,\tau)}) \\ &\quad \log \left(\frac{p(\mathbf{y}_{t'}^{(p,\tau)}, \mathbf{x}_t^{(m,\tau)})}{p(\mathbf{y}_{t'}^{(p,\tau)})p(\mathbf{x}_t^{(m,\tau)})} \right) d\mathbf{y}_{t'}^{(p,\tau)} d\mathbf{x}_t^{(m,\tau)}; \quad (2) \end{aligned}$$

par stationnarité des signaux X et Y , elle ne dépend que de $t' - t$.

L'auto-information mutuelle. Pour analyser la dynamique temporelle d'un processus X , on définit l'information mutuelle pour $Y = X$ et $t' - t = p\tau$.

$$I^{(m,p,\tau)}(X) = I(\mathbf{x}_{t+p\tau}^{(p,\tau)}, \mathbf{x}_t^{(m,\tau)}). \quad (3)$$

où la concaténation de $\mathbf{x}_{t+p\tau}^{(p,\tau)}$ et $\mathbf{x}_t^{(m,\tau)}$ correspond exactement au vecteur plongé dans un espace à $(m+p)$ -dimensions, $\mathbf{x}_t^{(m+p,\tau)}$:

$$\left(\mathbf{x}_{t+p\tau}^{(p,\tau)}, \mathbf{x}_t^{(m,\tau)} \right) = \mathbf{x}_t^{(m+p,\tau)}.$$

$I^{(m,p,\tau)}(X)$, que nous appelons l'auto-information mutuelle, mesure l'information partagée entre la dynamique de m points successifs et la dynamique des p points suivants. En utilisant cette définition, le taux d'entropie (eq. 1) peut-être réécrit [9] :

$$h^{(m,\tau)}(X) = H(X) - I^{(m,1,\tau)}(X). \quad (4)$$

Quand la distance τ séparant deux points consécutifs tend vers 0, la corrélation augmente et le taux entropie diminue. Réciproquement, quand $\tau \rightarrow +\infty$, les corrélations diminuent ainsi que l'information mutuelle, donc $h^{(m,\tau)}(X)$ augmente et sa valeur maximale est l'entropie du signal X .

Dans cette expression il y a deux contributions : $H(X)$, une quantité statique qui ne dépend que de la statistique à un point du signal X et $I^{(m,p,\tau)}(X)$, qui contient l'information sur la dynamique et qui est indépendante de la moyenne et de la variance de la distribution marginale de X [9]. Pour un processus multivarié stationnaire conjointement Gaussien, totalement défini par la donnée de sa variance σ^2 , et de sa fonction de corrélation normalisée $c(\tau)$ ($c(\tau = 0) = 1$), on a :

$$\begin{aligned} H^{(m,\tau)}(X) &= \frac{m}{2} \log(2\pi e\sigma^2) + \frac{1}{2} \log(|\Sigma^{(m)}|) \\ h^{(m,\tau)}(X) &= \frac{m}{2} \log(2\pi e\sigma^2) - \frac{1}{2} \log\left(\frac{|\Sigma^{(m)}|}{|\Sigma^{(m+1)}|}\right) \\ I^{(m,p,\tau)}(X) &= \frac{1}{2} \log\left(\frac{|\Sigma^{(m)}||\Sigma^{(p)}|}{|\Sigma^{(m+p)}|}\right), \end{aligned} \quad (5)$$

où $\Sigma^{(m)}$ représente la matrice de corrélation normalisée (de taille $m \times m$) du processus ($\Sigma^{(1)} = 1$).

2.2 Estimations

L'estimation du taux d'entropie se fait généralement à partir de l'équation (1) en utilisant l'estimateur d'entropie k-NN développé par Kozachenko et Leonenko [10], noté \hat{H} :

$$\hat{h}_1^{(m,\tau)}(X) = \hat{H}^{(m+1,\tau)}(X) - \hat{H}^{(m,\tau)}(X). \quad (6)$$

Nous proposons d'utiliser de préférence la définition (4) du taux d'entropie avec l'estimateur précédent pour l'entropie et l'estimateur de l'information mutuelle de Kraskov *et al* [11] qui repose aussi sur la recherche de voisins :

$$\hat{h}_2^{(m,\tau)}(X) = \hat{H}(X) - \hat{I}^{(m,1,\tau)}(X). \quad (7)$$

Les deux estimateurs, \hat{H} et \hat{I} , sont consistant avec une variance en $\mathcal{O}(1/N)$, et un biais en $\mathcal{O}(1/N^{\frac{1}{m}})$ pour l'estimateur d'entropie et en $\mathcal{O}(1/N^{\frac{1}{m+1}})$ pour l'estimateur de l'information mutuelle [12], où N est le nombre de points du signal analysé. La distance utilisée pour les deux estimateurs est la distance de Tchebychev.

3 Processus analysés

Nous étudions le biais et l'écart-type des deux estimateurs du taux d'entropie sur deux types de signaux de marginales et de fonctions de corrélation différentes mais de même taux d'entropie [13]. Le premier processus est un processus obtenu par

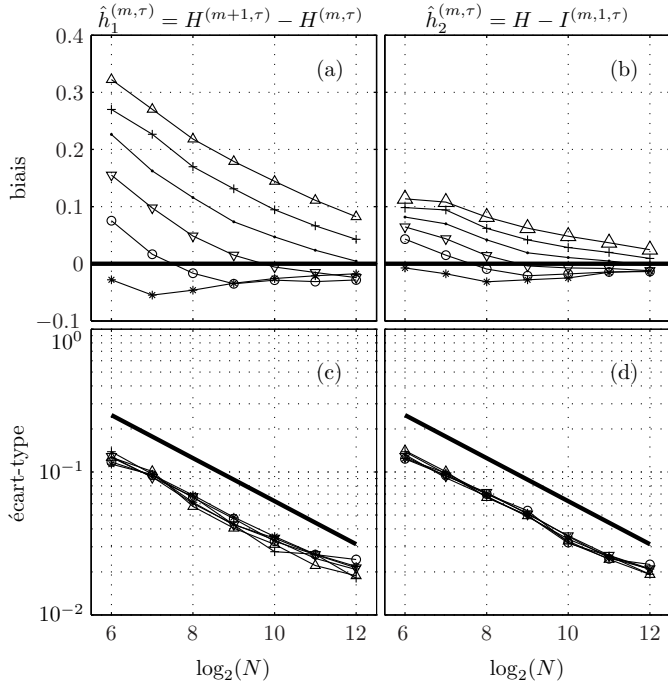


FIG. 1: **Performances à τ fixé.** Biases et écart type en fonction du nombre de points N pour les deux estimateurs du taux entropie. L'échelle τ est fixe ($\tau = 3$) et choisie de telle sorte que le coefficient de corrélation entre deux points successifs soit égal $c(\tau) = 0.94$.

filtrage passe-bas d'un bruit blanc Gaussien. Le filtre utilisé est de la forme :

$$F(\omega) = \frac{1}{1 + i \frac{f}{f_c}} \quad (8)$$

où f_c représente la fréquence de coupure du filtre et $i = \sqrt{-1}$. On normalise ensuite la variance du signal ($\sigma^2 = 1$) pour obtenir un signal Gaussien dont la fonction de corrélation s'écrit :

$$c(t - t') = e^{-2\pi f_c |t - t'|}. \quad (9)$$

Le taux d'entropie du signal $X_t^{(m,\tau)}$ se déduit des équations (5) en prenant les éléments de la matrice $\Sigma^{(m)}$ de la forme $\Sigma_{n,p}^{(m)} = c((p - n)\tau)$ pour n et p entier de 1 à m .

Le deuxième processus s'obtient en prenant l'exponentielle du processus Gaussien. On a alors un processus de marginale log-normale et de fonction de corrélation :

$$c_{LN}(t - t') = \frac{e^{c(t-t')} - 1}{e^1 - 1}. \quad (10)$$

Comme la moyenne du processus Gaussien est 0, l'entropie et le taux d'entropie du processus log-normal obtenu ont les même expressions (5) que le processus Gaussien associé [13].

4 Resultats

Dans cette section on compare les performances des deux estimateurs du taux l'entropie en fonction du paramètre d'échelle

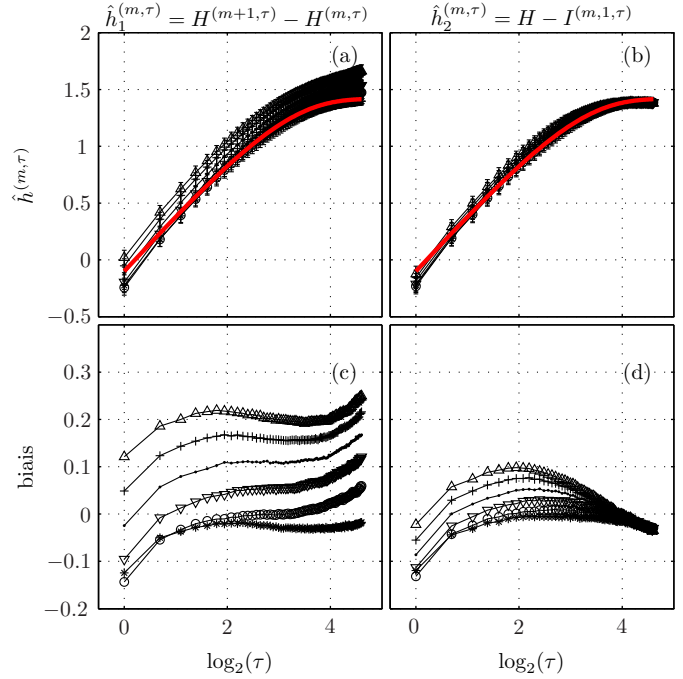


FIG. 2: **Dépendance en τ : cas Gaussien.** Valeurs obtenues et biais associés des deux estimateurs du taux d'entropie en fonction de l'échelle pour un processus Gaussien de taille $N = 2^{12}$. La courbe rouge représente les valeurs théoriques.

τ et de la dimension de plongement. Dans un premier temps, on fixe la valeur de τ et on varie la taille du signal. Dans un deuxième temps nous regarderons pour une taille fixé, le comportement du taux d'entropie en fonction de τ . Toutes les analyses sont faites en utilisant un nombre de voisin k fixé égal à 5. En effet, les différentes estimations ne dépendent que très légèrement de ce paramètre.

Dépendance en N . On génère le processus Gaussien et on choisit ensuite une valeur de τ telle que le coefficient de corrélation $c(\tau) = 0.94$. On calcule le taux d'entropie pour m variant de 1 à 6 et la taille du signal plongé $X_t^{(m,\tau)}$ variant de 2^6 à 2^{12} points. Le biais et l'écart-type des deux estimateurs sont présentés sur la figure 1. Quelque soit la dimension m , les écart-types obtenus sont très similaires et présentent un comportement en $\frac{1}{\sqrt{N}}$, représenté par la ligne continue. Par contre, le biais présente une grande variation en fonction de la dimension m avec des performances améliorées pour l'estimateur $\hat{h}_2^{(m,\tau)}$ basé sur l'information mutuelle. Les résultats obtenus pour les variables aléatoires log-normales associées sont similaires avec un biais deux fois plus important pour les deux estimateurs.

Dépendance en τ . Pour cette analyse on génère des processus de taille $N = 2^{12}$. On estime ensuite le taux d'entropie $h^{(m,\tau)}$ pour différentes valeurs de m et τ en utilisant les deux méthodes proposées. Pour chaque valeur de τ , nous construisons les τ différents plongements $X_t^{(m,\tau)}$ de taille N/τ possibles. Nous moyennons alors le taux d'entropie sur ces différentes versions. Ces opérations sont répétées sur 100 réali-

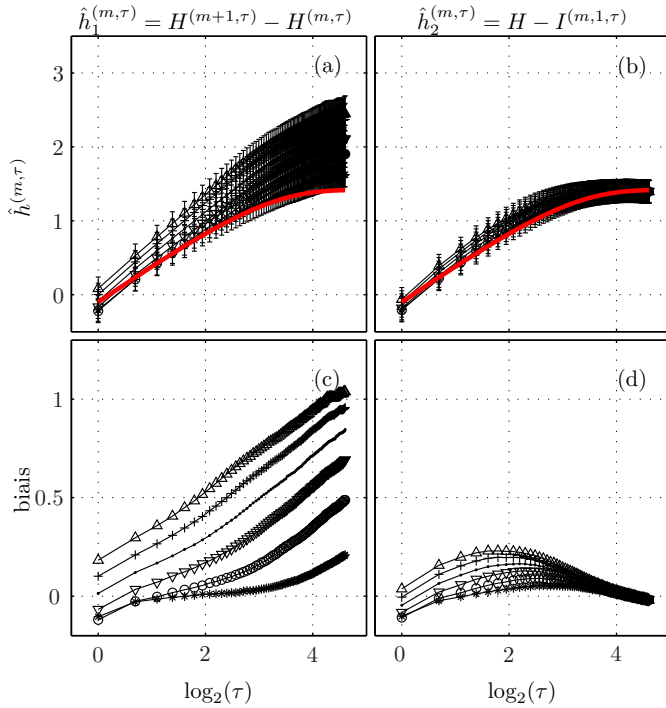


FIG. 3: **Dépendance en τ : cas log-normal.** Valeurs obtenues et biais associés des deux estimateurs du taux d'entropie à travers les échelles pour un processus log-normal de taille $N = 2^{12}$. La courbe rouge représente les valeurs théoriques.

sations différentes du processus. Il faut noter que quand l'échelle τ varie, non seulement l'estimation se fait sur des signaux de plus en plus petits (de taille N/τ), mais la corrélation entre deux coordonnées de $x_t^{(m,\tau)}$ diminue. Ainsi deux sources de biais possibles peuvent apparaître : la première due à la taille du signal étudié et l'autre due à la présence de corrélation dans le signal. Les résultats obtenus sur le processus Gaussien sont représentés dans la figure 2 et dans la figure 3 pour le processus log-normal. On voit clairement que le biais varie fortement avec la dimension m de plongement. Cette variation est beaucoup moins importante quand on utilise l'estimateur $\hat{h}_2^{(m,\tau)}$ basé sur l'information mutuelle. Quand le processus étudié n'est plus Gaussien (fig. 3), le biais de l'estimateur $\hat{h}_1^{(m,\tau)}$ augmente fortement et diverge clairement quand τ augmente et ceci même pour $m = 1$. Les performances de l'estimateur $\hat{h}_2^{(m,\tau)}$ sont clairement supérieures dans ce cas avec un biais qui ne diverge plus, quelque soit la dimension m . Ces courbes montrent ainsi que l'estimateur proposé dépend beaucoup moins de la taille et de la statistique du signal étudié et permet donc une meilleure estimation du taux d'entropie.

5 Conclusion et perspectives

Nous avons proposé l'utilisation du taux d'entropie afin de décrire la structure de corrélation de processus aléatoires à travers les échelles. Le taux d'entropie est en effet directement relié

à la fonction de corrélation si le processus a une statistique Gaussienne, mais contient dans le cas général plus d'informations car il dépend de toute la structure de corrélation, et non pas uniquement des corrélations à deux points.

Nous avons étudié l'évolution du taux d'entropie avec l'échelle de sous échantillonnage du signal, τ . L'estimateur utilisant la différence entre l'entropie et l'information mutuelle se révèle supérieur à celui défini directement par une différence d'entropies. En effet, il est non seulement moins biaisé et moins dépendant de la dimension de plongement utilisée mais son biais est aussi beaucoup moins dépendant de la statistique du signal, laquelle est, a priori, inconnue dans les applications.

References

- [1] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. XXVII, pp. 388–427, 1948.
- [2] R. X. Zhou, R. Cai, and G. Q. Tong, "Applications of entropy in finance: A review," *Entropy*, vol. 15, no. 11, pp. 4909–4931, Nov. 2013.
- [3] P. Abry, S. Roux, V. Chudáček, P. Borgnat, P. Gonçalves, and M. Doret, "Hurst exponent and intrapartum fetal heart rate: Impact of decelerations," in *26th International Symposium on Computer-Based Medical Systems (CBMS)*, 2013, pp. 1–6.
- [4] A. Kiani and M. R. Sahebi, "Edge detection based on the shannon entropy by piecewise thresholding on remote sensing images," *Iet Computer Vision*, vol. 9, no. 5, pp. 758–768, Oct. 2015.
- [5] S. M. Pincus and R. R. Viscarello, "Approximate entropy: a regularity measure for fetal heart rate analysis," *Obstet Gynecol*, vol. 79, no. 2, pp. 249–255, Feb 1992.
- [6] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039–H2049, Jun 2000.
- [7] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, p. 068102, Aug. 2002.
- [8] F. Takens, *Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, ch. Detecting strange attractors in turbulence, pp. 366–381. [Online]. Available: <http://dx.doi.org/10.1007/BFb0091924>
- [9] C. Granero-Belinchón, S. Roux, and N. Garnier, "Scaling of information in turbulence," *EPL*, vol. 115, p. 58003, 2016.
- [10] L. Kozachenko and N. Leonenko, "Sample estimate of entropy of a random vector," *Problems of Information Transmission*, vol. 23, pp. 95–100, 1987.
- [11] A. Kraskov, H. Stöbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, pp. 066138–1/066138–16, June 2004.
- [12] W. Gao, S. Oh, and P. Viswanath, "Demystifying fixed k-nearest neighbor information estimators," *CoRR*, vol. abs/1604.03006, 2016. [Online]. Available: <http://arxiv.org/abs/1604.03006>
- [13] K. Zografos and S. Nadarajah, "Expressions for renyi and shannon entropies for multivariate distributions," *Statistics & Probability Letters*, vol. 71, no. 1, pp. 71–84, Jan. 2005.