RÉNYI INFORMATION TRANSFER: PARTIAL RÉNYI TRANSFER ENTROPY AND PARTIAL RÉNYI MUTUAL INFORMATION

Septimia Sarbu

Department of Signal Processing Tampere University of Technology PO Box 527 FI-33101 Tampere, Finland septimia.sarbu@tut.fi

ABSTRACT

Shannon and Rényi information theory have been applied to coupling estimation in complex systems using time series of their dynamical states. By analysing how information is transferred between constituent parts of a complex system, it is possible to infer the coupling parameters of the system. To this end, we introduce the partial Rényi transfer entropy and we give an alternative derivation of the partial Rényi mutual information, using the conditional Rényi α -divergence. We prove that, in the limit $\alpha \rightarrow 1$, this divergence tends to the conditional Kullback-Leibler divergence from Shannon information theory. As a result, when $\alpha \rightarrow 1$, we obtain the partial transfer entropy and the partial mutual information from their Rényi equivalents. Using these Rényi information-theoretic functionals, we identify the coupling direction and delay between two processes in an autoregressive system of order 1.

Index Terms— partial Rényi transfer entropy, partial Rényi mutual information, information flow, coupling estimation

1. INTRODUCTION

Coupling estimation using time series from complex systems is an area of active research [1], [2], [3], [4]. In a complex system of interconnected parts, one can measure the dynamical states of each element of the system over time. But, there is no information directly available about how the elements are coupled in the system, what is the direction of the coupling, what is the strength of the coupling and what is the time delay of the coupling. Information theory is one approach to infer these parameters from time series of the dynamics of the system. Information theory analyses how information is processed by a dynamical system and how it is transmitted between systems. When two parts of a complex system are connected, they exchange information in one, two or both directions. Information-theoretic equations measure if the dynamical activities of parts of the complex system are statistically correlated. If they are, then there is a coupling between these parts. One can also determine the direction of the information flow, i.e. the direction of the coupling.

Two approaches from information theory have been used to analyse dynamical systems: Shannon information theory [5] and Rényi information theory [6]. In Shannon information theory, the transfer entropy(TE) [7] from one process Y to another process X measures how much shared information there is between the present state of X and the past states of Y, given that we know the history of the process X. It is an asymmetric measure of information, which makes it extremely useful in detecting directed exchange of information. The partial transfer entropy(PTE) [8] improves on TE, because it adds the environment as another conditioning variable to the transfer entropy. Two elements of a complex system, X and Y, might share information due to indirect connections through other parts of the system, Z. In estimating the transfer from Y to X, adding the knowledge about the processes Z to the knowledge of the history of X improves the estimation accuracy. The concept of eliminating indirect influences between two directly coupled processes X and Yis used to define the partial mutual information(PMI) in [9]. In order to make PMI a directed measure of information, the processes X and Y are time-delayed. In this case, the PMI has a maximum at the real coupling delay and has value close to zero at any other delays.

In Rényi information theory, the parameter α offers more flexibility than Shannon information theory, as it can be tuned according to the application. Moreover, the Rényi information-theoretic functionals include their Shannon equivalents as a special case when $\alpha \rightarrow 1$. The authors of [3] generalized TE to Rényi transfer entropy(RTE), by replacing the Shannon entropy in the definition of the TE with the Rényi entropy. They used RTE to measure the amount of information flow between the US, European and Asia-Pacific region markets. The authors of [10] generalized PMI to the conditional Rényi mutual information(CRMI).

In this paper, we extend the Rényi information theory to the partial Rényi transfer entropy and the partial Rényi mutual information. They are the generalization of PTE [8] and PMI [9], respectively. The paper is organized as follows: in section 2, we present the mathematical definitions and derivations of the partial Rényi transfer entropy and the partial Rényi mutual information, which are the core of this article. In section 3, we apply these equations to an autoregressive system of order 1, to identify the direction of the information flow. In section 4, we summarize the main points of this study.

2. MATHEMATICAL DEFINITIONS AND NOTATIONS

All the equations presented in this paper are derived for the discrete case. Let X, Y, Z be three random vectors and let x, y, z be a realization of these random vectors. Let p and q be two probability mass functions, with n elements, such that $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} q_i = 1$. Then, the Rényi α -divergence is defined in [6] as

$$D_{\alpha}(p \parallel q) = \frac{1}{\alpha - 1} \log \left(\sum_{i=1}^{n} \frac{p_i^{\alpha}}{q_i^{\alpha - 1}} \right) \tag{1}$$

In order to define the conditional Rényi α -divergence, we briefly review some facts about conditional random variables.

2.1. Conditional random variables

Let X and Y be two random variables(r.v.) and x and y one realization of these r.v.s, $x \in \mathcal{E}_X$ and $y \in \mathcal{E}_Y$, where \mathcal{E}_X is the ensemble of X and \mathcal{E}_Y is the ensemble of Y. Then, the conditional r.v. (X|Y = y) has the probability mass function(pmf) p(X|Y = y) = P((X|Y = y) = x) = P(X = x|Y = y), for every value $x \in \mathcal{E}_X$.

Let p and q be two pmfs of the r.v. (X|Y = y). Then, the Rényi α -divergence between p and q, $D_{\alpha}(p(X|Y = y) \parallel q(X|Y = y))$, is a function of y. As such, it is an r.v., because it is a function of the r.v. Y.

2.2. Rényi information-theoretic equations

In order to have a value for the divergence between two conditional pmfs, we define the conditional Rényi α -divergence as the expectation of $D_{\alpha}(p(X|Y = y) \parallel q(X|Y = y))$ with respect to the r.v. Y:

Definition 1.

$$D^*_{\alpha}(p(X|Y) \parallel q(X|Y)) =$$

= $\mathbb{E}_Y \left[D_{\alpha}(p(X|Y=y) \parallel q(X|Y=y)) \right]$
= $\frac{1}{\alpha - 1} \cdot \sum_y p(y) \cdot \log\left(\sum_x \frac{p^{\alpha}(x|y)}{q^{\alpha - 1}(x|y)}\right)$ (2)

This is an alternative definition to the one given in [10].

Theorem 1. With the above definitions and notations:

$$\lim_{\alpha \to 1} D^*_\alpha(p(X|Y) \parallel q(X|Y)) = D^*_{KL}(p(X|Y) \parallel q(X|Y)),$$

where D_{KL}^* is the conditional Kullback-Leibler divergence (the classical conditional divergence in Shannon information theory),

$$D_{KL}^{*}((p(X|Y) \parallel q(X|Y)) = \sum_{x} \sum_{y} p(x,y) \cdot \log \frac{p(x|y)}{q(x|y)}.$$

Proof.

$$x^{\alpha} = e^{\log x^{\alpha}} = e^{\alpha \cdot \log x}$$
$$\frac{\mathrm{d}}{\mathrm{d}\alpha} [x^{\alpha}] = \frac{\mathrm{d}}{\mathrm{d}\alpha} [e^{\alpha \cdot \log x}] = x^{\alpha} \cdot \log x$$

Using L'Hôspital's rule we have:

$$\lim_{\alpha \to 1} D_{\alpha}^{*}(p(X|Y) \parallel q(X|Y)) =$$

$$= \lim_{\alpha \to 1} \sum_{y} p(y) \cdot \frac{\mathrm{d}}{\mathrm{d}\alpha} \left[\log \left(\sum_{x} \frac{p^{\alpha}(x|y)}{q^{\alpha-1}(x|y)} \right) \right]$$

$$= \lim_{\alpha \to 1} \sum_{y} p(y) \cdot \frac{1}{\sum_{x} \frac{p^{\alpha}(x|y)}{q^{\alpha-1}(x|y)}} \sum_{x} q(x|y) \frac{\mathrm{d}}{\mathrm{d}\alpha} \left(\frac{p(x|y)}{q(x|y)} \right)^{\alpha}$$

$$= \lim_{\alpha \to 1} \sum_{y} p(y) \cdot \sum_{x} q(x|y) \cdot \left(\frac{p(x|y)}{q(x|y)} \right)^{\alpha} \cdot \log \frac{p(x|y)}{q(x|y)}$$

$$= \sum_{y} \sum_{x} p(y) \cdot p(x|y) \cdot \log \frac{p(x|y)}{q(x|y)}$$

$$= \sum_{x} \sum_{y} p(x, y) \cdot \log \frac{p(x|y)}{q(x|y)} = D_{KL}^{*}.$$
(3)

Definition 2. Let X, Y be two r.v.s. The Rényi mutual information is defined as the Rényi α -divergence between the joint pmf p(x, y) and the product of the marginals, $p(x) \cdot p(y)$:

$$RMI(X,Y) = D_{\alpha}(p(x,y) \parallel p(x) \cdot p(y))$$
$$= \frac{1}{\alpha - 1} \cdot \log\left(\sum_{x} \sum_{y} \frac{p(x,y)^{\alpha}}{(p(x) \cdot p(y))^{\alpha - 1}}\right)$$
(4)

Definition 3. Let X, Y, Z be three r.v.s. We define the conditional Rényi mutual information as the conditional Rényi α -divergence between the conditional joint pmf p(x, y|z) and the product of the conditional marginals, $p(x|z) \cdot p(y|z)$:

$$CRMI(X, Y|Z) = D^*_{\alpha}(p(x, y|z) \parallel p(x|z) \cdot p(y|z))$$
$$= \frac{1}{\alpha - 1} \cdot \sum_{z} p(z) \cdot \log\left(\sum_{x} \sum_{y} \frac{p(x, y|z)^{\alpha}}{(p(x|z) \cdot p(y|z))^{\alpha - 1}}\right)$$
(5)

Definition 4. We extend the concept of partial Shannon mutual information [9] to the Rényi information theory. We define the partial Rényi mutual information as a conditional Rényi mutual information between the r.v.s X, Y, Z:

$$PRMI(X, Y|Z) = CRMI(X, Y|Z) =$$

$$= \frac{1}{\alpha - 1} \cdot \sum_{z} p(z) \cdot \log\left(\sum_{x} \sum_{y} \frac{p^{\alpha}(x, y|z)}{(p(x|z) \cdot p(y|z))^{\alpha - 1}}\right)$$
(6)

The concept of Rényi transfer entropy was introduced in [3]. Here, we use an alternative definition based on the conditional Rényi mutual information.

Definition 5. Let X, Y be two random vectors, $X = [X_{n+1}X_n \cdots X_{n-k+1}]$, $Y = [Y_nY_{n-1}\cdots Y_{n-l+1}]$, where the current time point is n + 1 and k, l are the time lags for X and Y respectively. Then,

$$RTE_{Y->X}(k,l) = CRMI(X_{n+1}, [Y_nY_{n-1}\cdots Y_{n-l+1}] | [X_nX_{n-1}\cdots X_{n-k+1}]) =$$

$$= \frac{1}{\alpha - 1} \sum_{x_n} \cdots \sum_{x_{n-k+1}} p(x_n, \cdots, x_{n-k+1}) \cdot \log \sum_{x_{n+1}} \sum_{y_n} \sum_{y_{n-l+1}} \frac{p^{\alpha}(x_{n+1}, y_n, \cdots, y_{n-l+1} | x_n, \cdots, x_{n-k+1}))}{p^{\alpha - 1}(x_{n+1} | x_n, \cdots, x_{n-k+1})}$$

$$\overline{p^{\alpha - 1}(y_n, \cdots, y_{n-l+1} | x_n, \cdots, x_{n-k+1})}$$
(7)

Definition 6. We introduce the Partial Rényi Transfer Entropy. Let X, Y, Z be three random vectors, $X = [X_{n+1}X_n \cdots X_{n-k+1}]$, $Y = [Y_nY_{n-1}\cdots Y_{n-l+1}]$ and $Z = [Z_nZ_{n-1}\cdots Z_{n-m+1}]$, where the current time point is n + 1 and k, l, m are the time lags for X, Y and Z respectively. For clarity, let $w = [x_n, \cdots, x_{n-k+1}, z_n, \cdots, z_{n-m+1}]$. Then,

$$PRTE_{Y->X}(k,l,m) = CRMI(X_{n+1}, [Y_nY_{n-1}\cdots Y_{n-l+1}] | [X_nX_{n-1}\cdots X_{n-k+1}], [Z_nZ_{n-1}\cdots Z_{n-m+1}]) = \frac{1}{\alpha-1} \sum_{w} p(w) \cdot \log \sum_{x_{n+1}} \sum_{y_n} \cdots \sum_{y_{n-l+1}} \frac{p^{\alpha}(x_{n+1}, y_n, \cdots, y_{n-l+1})|w)}{[p(x_{n+1}|w) \cdot p(y_n, \cdots, y_{n-l+1}|w)]^{\alpha-1}}$$
(8)

3. SIMULATION RESULTS

We used the Rényi information-theoretic equations to find the direction of information transfer in an autoregressive(AR) system of order 1. The system is composed of three coupled processes:

$$\begin{cases} X[n] = 0.6 \cdot X[n-1] + \epsilon_1 \\ Y[n] = 0.9 \cdot Y[n-1] + X[n-1] + \epsilon_2 \\ Z[n] = 0.2 \cdot Z[n-1] + 0.5 \cdot Y[n-1] + X[n-1] + \epsilon_3 \end{cases}$$

The processes X[n] and Y[n] are coupled in the direction from X[n] to Y[n], with the coupling delay equal to 1 and the coupling strength equal to 1. The processes Y[n] and Z[n] are coupled in the direction from Y[n] to Z[n], with the coupling delay equal to 1 and the coupling strength equal to 0.5. The processes X[n] and Z[n] are coupled in the direction from X[n] to Z[n], with the coupling delay equal to 1 and the coupling strength equal to 1. We added to the processes Gaussian noise distributed as $\epsilon_1, \epsilon_2, \epsilon_3 \sim \mathcal{N}(0, 10^{-6})$.

In the definition of the Rényi information-theoretic equations, we used time lags equal to k = l = m = 1. We set the value of the parameter $\alpha = 3$. We generated time series [x(n)], [y(n)], [z(n)], using $N_t = 50$ time points for each time series. The system starts from an initial state $[x_0 \ y_0 \ z_0] = [1 + \mathcal{U}(0, 1) \ 1 + \mathcal{U}(0, 1)], 1 + \mathcal{U}(0, 1)],$ where $\mathcal{U}(0, 1)$ is the standard uniform distribution. The results are averaged over 100 simulations.

In order to make the computations more straightforward, we transformed the conditional pmfs into joint pmfs, using the identity $p(x|y) = \frac{p(x,y)}{p(y)}$ from probability theory. As a result, the Rényi functionals that we used to detect the coupling in the AR system became:

$$PRTE_{Y->X|Z}(1,1,1) = \frac{1}{\alpha-1} \sum_{x_n} \sum_{z_n} p(x_n, z_n) \cdot \\ \cdot \log \sum_{x_{n+1}} \sum_{y_n} \frac{p^{\alpha}(x_{n+1}, y_n | x_n, z_n)}{\left[p(x_{n+1} | x_n, z_n) \cdot p(y_n | x_n, z_n) \right]^{\alpha-1}} = \frac{1}{\alpha-1} \cdot \\ \cdot \sum_{x_n} \sum_{z_n} p(x_n, z_n) \cdot \log \sum_{x_{n+1}} \sum_{y_n} \frac{p^{\alpha}(x_{n+1}, y_n, x_n, z_n)}{p^{\alpha}(x_n, z_n)} \cdot \\ \cdot \frac{p^{\alpha-1}(x_n, z_n)}{p^{\alpha-1}(x_{n+1}, x_n, z_n)} \cdot \frac{p^{\alpha-1}(x_n, z_n)}{p^{\alpha-1}(y_n, x_n, z_n)} = \\ = \frac{1}{\alpha-1} \sum_{x_n} \sum_{z_n} p(x_n, z_n) \cdot \\ \cdot \log \sum_{x_{n+1}} \sum_{y_n} \frac{p^{\alpha}(x_{n+1}, y_n, x_n, z_n) \cdot p^{\alpha-2}(x_n, z_n)}{\left[p(x_{n+1}, x_n, z_n) \cdot p(y_n, x_n, z_n) \right]^{\alpha-1}}$$
(9)

Similarly:

$$PRMI(X, Y|Z) = \frac{1}{\alpha - 1} \sum_{z} p(z) \cdot \log \sum_{x} \sum_{y}$$

$$\frac{p^{\alpha}(x, y, z) \cdot p^{\alpha - 2}(z)}{[p(x, z) \cdot p(y, z)]^{\alpha - 1}}$$

$$RTE_{Y - > X}(1, 1) = \frac{1}{\alpha - 1} \sum_{x_n} p(x_n) \cdot \log \sum_{x_{n+1}} \sum_{y_n}$$
(10)

$$\frac{p^{\alpha}(x_{n+1}, y_n, x_n) \cdot p^{\alpha-2}(x_n)}{[p(x_{n+1}, x_n) \cdot p(y_n, x_n)]^{\alpha-1}}$$
(11)

$$RMI(X,Y) = \frac{1}{\alpha - 1} \log \sum_{x} \sum_{y} \frac{p^{\alpha}(x,y)}{[p(x) \cdot p(y)]^{\alpha - 1}} \quad (12)$$

We estimated the joint pmfs using multivariate kernel density estimation implemented as the Matlab toolbox "Kernel Density Estimation Toolbox for Matlab" [11]. We estimated two, three and four-dimensional pmfs. We computed the marginal pmfs from the joint pmf by summing out the dimensions that were not of interest. We pluged-in the values of the estimated joint and marginal pmfs, to obtain the estimate of the Rényi information-theoretic equations.

Given the time series from the AR system, the objective is to find the direction of the coupling and the time delay of the coupling. Nothing is known about the structure of the system, i.e. the value of the parameters, the coupling strength, the coupling delay, the distribution of the noise and how many terms there are in the equation of each process. For this purpose, we measured the information transfer between the processes Y and Z in both directions. We computed the difference between the information flow from Y to Z and the one from Z to Y, where τ is the time delay, i.e.

$$\begin{split} &\Delta RMI = RMI(Y[n], Z[n+\tau]) - RMI(Z[n], Y[n+\tau]) \\ &\Delta RTE = RTE_{Y->Z}(1,1) - RTE_{Z->Y}(1,1) \\ &\Delta PRMI = PRMI(Y[n], Z[n+\tau]|X) - \\ &- PRMI(Z[n], Y[n+\tau]|X) \\ &\Delta PRTE = PRTE_{Y->Z|X}(1,1) - PRTE_{Z->Y|X}(1,1). \end{split}$$

We computed these measures for several coupling time delays, from $\tau = 1$ to $\tau = 10$. The results are shown in the following table. If these measures are positive, then the processes Y and Z are coupled and the information flow is from Y to Z. If they are negative, then the connection is from Z to Y. $\Delta PRMI$ and $\Delta PRTE$ are positive for all time delays, which shows that the connection is from Y to Z. In addition, they correctly identify the coupling delay at $\tau = 1$, because their maximum value is at this time delay. They exhibit a deacreasing trend as the coupling delay increases to $\tau = 10$. This indicates that, as the time delay increases, the processes are less and less correlated.

Similarly to the other measures, ΔRTE has a decreasing trend with the time delay. But, it fails to identify the correct coupling delay, because its maximum value is at $\tau = 3$ and it has almost identical values for time delays 1, 2, 3. Moreover, at time delays $\tau = 9$ and $\tau = 10$, the ΔRTE is negative, meaning that it does not capture the directed link from Y to Z. ΔRMI is negative for all time delays, which shows that it cannot find the real direction of information propagation.

4. CONCLUSIONS

In this study, we used a version of the Rényi conditional α divergence to introduce the partial Rényi transfer entropy and to derive an alternative definition of the partial Rényi mutual information. We proved a theorem that states that this divergence tends to the classical conditional Kullback-Leibler divergence, as the Rényi α parameter tends to 1. As a result,

Table 1. Measures of Rényi information transfer for various coupling time delays

Time	$\Delta \mathrm{RMI}$	ΔRTE	$\Delta \mathrm{PRMI}$	$\Delta PRTE$
delay				
$\tau = 1$	-0.299	4.432	7.356	3.57
$\tau = 2$	-0.22	4.431	7.18	3.37
$\tau = 3$	-0.358	4.475	6.933	3.299
$\tau = 4$	-0.479	4.229	6.669	3.272
$\tau = 5$	-0.6	3.913	6.348	3.413
$\tau = 6$	-0.7	3.214	6.132	3.059
$\tau = 7$	-0.813	2.305	5.845	2.748
$\tau = 8$	-0.891	0.564	5.629	2.123
$\tau = 9$	-0.941	-1.275	5.315	1.915
$\tau = 10$	-0.98	-3.023	5.005	1.47

all the Rényi information-theoretic functionals defined using this divergence measure will tend to their Shannon counterparts, as the Rényi α parameter tends to 1. As an example, we showed that the partial Rényi transfer entropy and the partial Rényi mutual information correctly identify the direction of the information flow and the coupling delay between two coupled processes in an AR system of order 1.

As future work, we will investigate other types of estimates than the plug-in estimates used in this article and analyse their statistical properties, as studied by the authors of [10]. An overview of other available estimation methods for information-theoretic functionals can be found in [1]. We will also apply these Rényi information transfer measures to other benchmark simulated systems: discrete-time systems, such as the tent map, the Hénon map and the logistic map, continuous-time systems, such as coupled Lorenz systems and coupled Rössler-Lorenz systems, as well as to time series from real complex systems. In addition, we will investigate the influence of the Rényi α parameter on the accuracy of the inference of information flow in these test systems.

5. ACKNOWLEDGEMENTS

The work presented in this paper was funded by the Graduate School in Electronics, Telecommunication and Automation(GETA), Aalto University, Helsinki, and the Academy of Finland project number 310071.

6. REFERENCES

 K. Hlaváčková, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on informationtheoretic approaches in time series analysis," *Physics Reports*, vol. 441, pp. 1–46, 2007.

- [2] Papana A. and Kugiumtzis D., "Detection of directionality of information transfer in nonlinear dynamical systems," in *Topics on Chaotic Systems, Selected papers* from Chaos 2008 International Conference, World Scientific, 2009, pp. 251–264.
- [3] P. Jizba, H. Kleinert, and M. Shefaat, "Rényi's information transfer between financial time series," *Physica A*, vol. 391, pp. 2971–2989, 2012.
- [4] D. Kugiumtzis, "Partial transfer entropy on rank vectors," *The European Physical Journal Special Topics*, vol. 222, pp. 401–420, 2013.
- [5] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379– 423, 1948.
- [6] Rényi A., "On measures of entropy and information," in Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, 1961, vol. 1, pp. 547–561.
- [7] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, pp. 461–464, 2000.
- [8] V.A. Vakorin, O.A. Krakovska, and A.R. McIntosh, "Confounding effects of indirect connections on causality estimation," *Journal of Neuroscience Methods*, vol. 184, pp. 152–160, 2009.
- [9] S. Frenzel and B. Pompe, "Partial mutual information for coupling analysis of multivariate time series," *Physical Review Letters*, vol. 99, pp. 204101 1–4, November 2007.
- [10] B. Poczos and J. Schneider, "Nonparametric estimation of conditional information and divergences," in *International Conference on AI and Statistics (AISTATS)*, 2012, vol. XX, pp. 914–923.
- [11] Ihler A. and Mandel M., "Kernel density estimation toolbox for matlab," in *http://www.ics.uci.edu/~ihler/code/kde.html*.