ECEA 2017 4th International Electronic Conference on Entropy and Its Applications 21 November–1 December 2017

sciforum

Conference Proceedings Paper

On a General Definition of Conditional Rényi Entropies

Velimir M. Ilić [1,*], Ivan B. Djordjević [2] and Miomir Stanković [3]

1 Mathematical Institute of the Serbian Academy of Sciences and Arts, Kneza Mihaila 36, 11000 Beograd, Serbia
2 Department of Electrical and Computer Engineering, University of Arizona, 1230 E. Speedway Blvd, Tucson, AZ, USA
3 Faculty of Occupational Safety, University of Niš, Čarnojevića 10a, 18000 Niš, Serbia
* Correspondence: velimir.ilic@gmail.com

Abstract: In recent decades, different definitions of conditional Rényi entropy (CRE) have been introduced. Thus, Arimoto proposed a definition that found an application in information theory, Jizba and Arimitsu proposed a definition that found an application in time series analysis and Renner-Wolf, Hayashi and Cachin proposed definitions that are suitable for cryptographic applications. However, there is still no a commonly accepted definition, nor a general treatment of the CRE-s, which can essentially and intuitively be represented as an average uncertainty about a random variable X if a random variable Y is given. In this paper we fill the gap and propose a three-parameter CRE, which contains all of the previous definitions as special cases that can be obtained by a proper choice of the parameters. Moreover, it satisfies all of the properties that are simultaneously satisfied by the previous definitions, so that it can successfully be used in aforementioned applications. Thus, we show that the proposed CRE is positive, continuous, symmetric, permutation invariant, equal to Rényi entropy for independent X and Y, equal to zero for $X = Y$ and monotonic. In addition, as an example for the further usage, we discuss the properties of generalized mutual information, which is defined using proposed CRE.

Keywords: Rényi entropy; conditional entropy; mutual information

PACS: 89.70.Cf, 87.19.lo

1. Introduction

Rényi entropy (RE) is a well known one parameter generalization of Shannon entropy. It has successfully been used in a number of different fields, such as statistical physics, quantum mechanics, communication theory and data processing [1,2]. On the other hand, the generalization of conditional Shannon entropy to Rényi entropy case is not uniquely defined. Thus, different definitions of conditional Rényi entropy has been proposed in the context of channel coding [3], secure communication [4–6] and multifractal analysis [1]. However, no one of the generalization satisfies a set of basic properties which are satisfied by the Shannon conditional entropy and there is no general agreement about the proper definition, so the choice of the definition depends on application purpose. Essentially, in all of the previous discussions, CRE can be represented as an average uncertainty about a random variable X if a random variable Y is given.

In this paper, we introduce three-parameter CRE which contains previously defined conditional entropies as special cases that can be obtained by a proper choice of the parameters. Moreover, it satisfies all of the properties that are simultaneously satisfied by the previous definitions, so that it can successfully be used in aforementioned applications. Thus, we show that the proposed CRE is positive,

continuous, symmetric, permutation invariant, equal to Rényi entropy for independent X and Y, equal to zero for $X = Y$ and monotonic.

One of the most frequent usages of conditional entropies is for the definition of mutual information (MI). The MI represents information transfer between the channel input and output which can be defined as the input uncertainty reduction if the output symbols are known. Thus, the MI which corresponds to the α-β-γ entropy measured as a difference between the input RE and the input α-β-γ CRE when the output is given. We analyze the properties of α-β-γ MI and show that the basic properties of Shannon MI, such as continuity, annulation for independent input and output and reducing to the output entropy for independent events, are also satisfied in the case of α-β-γ MI, which further validates the usage of α-β-γ CRE.

The paper is organized as follows. In Section 2 we review basic notions about Rényi entropy. The definition of α-β-γ CRE is introduced in Section 3 and its properties are considered in Section 4. The α-β-γ MI is considered in Section 5.

2. Rényi Entropy

Let X be a discrete random variable taking values from a sample space $\{x_1, \ldots, x_n\}$ and distributed according to $P_X = (p_1, \ldots, p_n)$. The Rényi entropy of X of order α ($\alpha > 0$) is defined as

$$\mathcal{R}_\alpha(X) = \frac{1}{1-\alpha} \log_2 \left(\sum_x P_X(x)^\alpha \right). \tag{1}$$

For two discrete random variables X and Y, with joint density P_{XY}, the joint Rényi entropy is defined with

$$\mathcal{R}_\alpha(X, Y) = \frac{1}{1-\alpha} \log_2 \left(\sum_{x,y} P_{XY}(x, y)^\alpha \right). \tag{2}$$

The following properties hold [1]:

(A1) $\mathcal{R}_\alpha(X)$ is continuous with respect to P_X;
(A2) Adding a zero probability event to the sample space of X does not change $\mathcal{R}_\alpha(X)$
(A3) $\mathcal{R}_\alpha(X)$ takes its largest value for the uniformly distributed random variable i.e.,

$$\mathcal{R}_\alpha(X) \leq \log_2 n, \tag{3}$$

with equality iff $P_X = (1/n, \ldots, 1/n)$.
(A4) If $P_X = (1, 0, \ldots, 0)$, then $\mathcal{R}_\alpha(X) = 0$;
(A5) $\mathcal{R}_\alpha(X)$ is symmetric with respect to P_X, i.e., if $X \sim (p_1, \ldots, p_n)$ and $Y \sim ((p_{\pi(1)}, \ldots, p_{\pi(n)}))$, where π is any permutation of $\{1, \ldots, n\}$, than $\mathcal{R}_\alpha(X) = \mathcal{R}_\alpha(Y)$ for all P_X;
(A6) $\mathcal{R}_\alpha(X)$ is continuous with respect to α and in the limit case reduces to Shannon entropy $\mathcal{S}(X)$ [7]:

$$\mathcal{S}(X) = \lim_{\alpha \to 1} \mathcal{R}_\alpha(X) = -\sum_x P_X(x) \log_2 P_X(x); \tag{4}$$

(A7) Rényi entropy can be represented as quasi-linear mean of Hartley information content $H = -\log_2 P_X$ as:

$$\mathcal{R}_\alpha(X) = g_\alpha^{-1} \left(\sum_x P(x) g_\alpha(H(x)) \right) \tag{5}$$

where the function g_α is defined with

$$g_\alpha(x) = \begin{cases} \dfrac{2^{(1-\alpha)x-1}}{1-\alpha}, & \text{for} \quad \alpha \neq 1; \\ x, & \text{for} \quad \alpha = 1. \end{cases} \tag{6}$$

or any linear function of Equation (6) (it follows from a well known result from mean value theory: if one function is a linear function of another one, they generate the same quasi-linear mean).

3. α-β-γ Conditional Rényi Entropy

Previously, several definitions of the CRE has been proposed [8] as a measure of average uncertainty about random variable Y when X is known. In this section, we unify all of this definitions and we define the CRE as three parameter function which can access all of the previous definitions by a special choice of the parameters.

Let $(X,Y) \sim P_{X,Y}$, $X \sim P_X$. The Rényi entropy of conditional random variables $Y|X=x$ distributed according to $P_{Y|X=x}$ is denoted with $\mathcal{R}_\alpha(Y|X=x)$. The conditional Rényi entropy is defined with

$$R_\alpha^{\beta,\gamma}(Y|X) = g_\gamma^{-1}\left(\sum_x P_X^{(\beta)}(x) g_\gamma(\mathcal{R}_\alpha(Y|X=x))\right) \tag{7}$$

where g_γ is given with Equation (6) and escort distribution of P_X is defined with:

$$P_X^{(\beta)}(x) = \frac{P_X(x)^\beta}{\sum_x P_X(x)^\beta}. \tag{8}$$

The definition can straightforwardly be extended to the joint conditional entropy. For random variables X,Y,Z, we define joint conditional entropy as:

$$R_\alpha^{\beta,\gamma}(Y,Z|X) = g_\gamma^{-1}\left(\sum_x P_X^{(\beta)}(x) g_\gamma(\mathcal{R}_\alpha(Z,Y|X=x))\right) \tag{9}$$

The definitions extends to the case $\beta=\infty$, by taking a limit $\beta\to\infty$, and by using $\lim_{\beta\to\infty} P_X^{(\beta)}(x) = \max_x P_X(x)$. In the case of $\alpha=\beta=1$, the definitions reduces to the Shannon case. By choosing appropriate values for β and γ we get the previously considered definitions as follows:

[C-CRE] $\beta=\gamma=1$, Cachin [4]

$$R_\alpha^C(Y|X) = \sum_x P_X(x)\ \mathcal{R}_\alpha(Y|X=x) \tag{10}$$

[JA-CRE] $\beta=\gamma=\alpha$, Jizba and Arimitsu [1]

$$R_\alpha^{JA}(Y|X) = \frac{1}{1-\alpha}\log_2 \sum_x P_X^{(\alpha)}(x) 2^{(1-\alpha)\mathcal{R}_\alpha(Y|X=x)} \tag{11}$$

[RW-CRE] $\beta=\infty$ Renner and Wolf [5] (RW-CRE)

$$R_\alpha^{RW}(Y|X) = \begin{cases} \min_x R_\alpha(Y|X=x) & \text{if} \quad \alpha>1 \\ \max_x R_\alpha(Y|X=x) & \text{if} \quad \alpha<1 \end{cases} \tag{12}$$

[A-CRE] $\beta=1, \gamma=2-\alpha^{-1}$ Arimoto [3] (A-CRE)

$$R_\alpha^{A}(Y|X) = \frac{1}{1-\alpha}\log_2 \sum_x P_X(x) 2^{\frac{1-\alpha}{\alpha}\mathcal{R}_\alpha(Y|X=x)} \tag{13}$$

[H-CRE] $\alpha=\gamma, \beta=1$ Hayashi [6] (H-CRE).

$$R_\alpha^{H}(Y|X) = \frac{1}{1-\alpha}\log_2 \sum_x P_X(x) 2^{(1-\alpha)\mathcal{R}_\alpha(Y|X=x)} \tag{14}$$

4. Properties of α-β-γ CRE

The α-β-γ CRE satisfies the set of important properties for all α, β, γ:

(B1) $R_\alpha^{\beta,\gamma}(Y|X) \geq 0$
(B2) $R_\alpha^{\beta,\gamma}(Y|X)$ is continuous with respect to $P_{X,Y}$;
(B3) $R_\alpha^{\beta,\gamma}(Y|X)$ is symmetric with respect to $P_{Y|X=x}$ for all $P_X \in \Delta_n$, $i = 1, \ldots, n$ and $P_{Y|X=x}$;
(B4) If $P_{Y|X=x}$ is a permutation of $P_{Y|X=x_1}$, for all x, then, and $R_\alpha^{\beta,\gamma}(Y|X) = \mathcal{R}(Y|X = x_1)$
(B5) If X and Y are independent, then $R_\alpha^{\beta,\gamma}(Y|X) = \mathcal{R}(Y)$
(B6) If $X = Y$, then $R_\alpha^{\beta,\gamma}(Y|X) = 0$
(B7) In the case of $\alpha = \beta = 1$, the definitions reduces to the Shannon case.
(B8) Let X, Y, Z be random variables distributed according with joint distribution $P_{X,Y,Z}$ and corresponding marginal distributions $P_X, P_Y, P_{Y,Z}$

$$R_\alpha^{\beta,\gamma}(Y, Z|X) \leq R_\alpha^{\beta,\gamma}(Y|X) \tag{15}$$

The proofs for the properties B1–B6 straightforwardly follows from the definition Equation (7) of the conditional Rényi entropy and from the properties A1–A5 of Rényi entropy. Here, we give the proof for the property B8.

Proof of the Property 8: Let, $P_{Y,Z|x}$, $P_{Z|x}$, $P_{Y|x,z}$ be conditional distributions. Similarly as in [9], we have:

$$\begin{aligned} \sum_{y,z} P_{Y,Z|x}(y,z)^\alpha &= \sum_z P_{Z|x}(z)^\alpha \sum_y P_{Y|x,z}(y)^\alpha \\ &\begin{cases} \leq \sum_z P_{Z|x}(z)^\alpha, & \text{for} \quad \alpha > 1 \\ \geq \sum_z P_{Z|x}(z)^\alpha, & \text{for} \quad \alpha < 1 \end{cases} \end{aligned} \tag{16}$$

so that we have $\mathcal{R}_\alpha(Y, Z|X = x) \leq \mathcal{R}_\alpha(Y|X = x)$ and result follows from the definition of conditional entropy since g_γ is increasing. □

Additional properties which are satisfied in the case of Shannon entropy are not satisfied in general and are limited to a special choices of β and γ.

(B9) Chain rule:

$$\begin{aligned} \mathcal{R}_\alpha(X, Y) &= \mathcal{R}_\alpha(X) + R_\alpha^{\beta,\gamma}(Y|X) \\ &= \mathcal{R}_\alpha(Y) + R_\alpha^{\beta,\gamma}(X|Y) \end{aligned} \tag{17}$$

is satisfied in general only in the case of JA-CRE. Jizba and Arimitsu [1] used it (with an assumption that g_γ is invertible and positive) as one of the Generalized Shannon-Khinchin axioms, along with the properties A1-A3 of Rényi entropy, and shown that Réni can be characterized as a unique function which satisfies them. It also implies the symmetry of generalized mutual information introduced in the next section.

(B10) Weak chain rule:

$$R_\alpha^{\beta,\gamma}(Y|X) \geq \mathcal{R}_\alpha(X, Y) - \log_2 m \tag{18}$$

(B11) Conditioning reduces the entropy (CRE),

$$R_\alpha^{\beta,\gamma}(Y|X) \leq \mathcal{R}_\alpha(X) \tag{19}$$

is satisfied in the cases of H-CRE, A-CRE and RW-CRE (for $\alpha \geq 1$). CRE states that an additional knowledge can not increase the information. Although it can intuitively be treated

as an ineluctable property, breaking the CRE can still be interpreted using concept of spoiling knowledge as in [4].

(B12) Monotonicity says that if X, Y, and Z forms Markov chain then

$$R_\alpha^{\beta,\gamma}(X|Z) \leq R_\alpha^{\beta,\gamma}(X|Y) \tag{20}$$

It holds in the case of A-CRE and H-CRE and implies data processing inequality (defined in the next section), which is an important property for applications of Rényi entropy in cryptography [5].

The previous discussion is summarized in Table 1 [9].

Table 1. Properties of different CRE-s (✓ stands for satisfied, ✗ for not satisfied, and ✓* for satisfied for $\alpha \geq 1$).

	C	*JA*	*RW*	*A*	*H*
Chain Rule	✗	✓	✗	✗	✗
Weak Chain Rule	✗	✓	✗	✓	✗
CRE	✗	✗	✓*	✓	✓
Monotonicity	✗	✓	✗	✓	✓

Since the additional information theoretic properties are not satisfied in general, no one of the previous definitions has not been commonly accepted, and the choice of the definition depends on application purpose. On the other hand, a set of properties B1–B8 which are satisfied for all of them, also hold in case of α-β-γ CRE defined by Equation (7), which justifies the definition.

5. α-β-γ Mutual Information

A communication channel with input X and output Y described by transition matrix $P_{Y|X}$

$$P_{Y|X}^{(j,i)} = P_{Y|X=x_i}(Y = y_j|X = x_i)$$

The mutual information is a measure of information transfer between input and output of communication channel. Thus, if the uncertainty about the input is measured by $\mathcal{R}_\alpha(X)$ and its uncertainty after all symbols are received by $\mathcal{R}_\alpha^{(\beta,\gamma)}(X|Y)$, the α-β-γ mutual information between X and Y is defined as

$$I_\alpha^{\beta,\gamma}(X,Y) = \mathcal{R}_\alpha(X) - R_\alpha^{\beta,\gamma}(X|Y). \tag{21}$$

This definition generalizes the previously introduced one by Arimoto who used A-CRE, and one by Jizba et al. [10] who used JA-CRE.

By usage of the properties of α-β-γ CRE, it is easy to conclude that the basic properties are satisfied for all α, β, γ:

(D1) $I_\alpha^{\beta,\gamma}(Y,X)$ is continuous with respect to $P_{X,Y}$;
(D2) If X and Y are independent, then $I_\alpha^{\beta,\gamma}(X,Y) = 0$
(D3) If $X = Y$, then $I_\alpha^{\beta,\gamma}(X,Y) = \mathcal{R}_\alpha(X)$

However, another properties which hold in the Shannon case are not satisfied in general:

(D4) Non-Negativity $I_\alpha^{\beta,\gamma}(Y,X) \geq 0$
(D5) Symmetry: $I_\alpha^{\beta,\gamma}(X,Y) = I_\alpha^{\beta,\gamma}(Y,X)$
(D6) Data processing inequality (DPI): If X, Y, and Z forms Markov chain then $I_\alpha^{\beta,\gamma}(X,Y) \geq I_\alpha^{\beta,\gamma}(X,Z)$

In Table 2 we list the properties and their fulfilments for the cases when the mutual information is defined using C, JA, RW, A and H conditional Rényi entropies [9]. Thus, if the MI is defined using any of the previously considered CRE definitions, it fails down to satisfy some of the properties properties D4-D6. On the other hand, the set of properties D1-D3 which is sattisfied for all of them is also satisfied for α-β-γ MI, which justify its usage as a measure of information transfer.

Table 2. Properties of different MI deffinitions (✓ stands for satisfied, ✗ for not satisfied, and ✓* for satisfied for $\alpha \geq 1$).

	C	*JA*	*RW*	*A*	*H*
Non-Negativity	✗	✗	✓*	✓	✓
Symmetry	✗	✓	✗	✗	✗
DPI	✗	✓	✗	✓	✓

6. Conclusions

We introduced α-β-γ conditional Rényi entropy (CRE) which contains previously defined conditional entropies as special cases that can be obtained by a proper choice of the parameters [1,3–6]. It satisfies all of the properties that are simultaneously satisfied by the previous definitions, so that it could be successfully be used in channel coding, secure communication and multifractal analysis.

In addition, we analyzed the properties of mutual information (MI) which is defined using α-β-γ CRE. The resulting MI measure satisfies the set of basic properties which further validates the usage of α-β-γ CRE.

Acknowledgments: Research supported by Ministry of Science and Technological Development, Republic of Serbia, Grants Nos. ON 174026 and III 044006.

References

1. Jizba, P.; Arimitsu, T. The world according to Rényi: thermodynamics of multifractal systems. *Annals of Physics* **2004**, *312*, 17 – 59.
2. Csiszár, I. Generalized cutoff rates and Renyi's information measures. *Information Theory, IEEE Transactions on* **1995**, *41*, 26–34.
3. Arimoto, S. Information Mesures and Capacity of Order α for Discrete Memoryless Channels. Topics in Information Theory; Csiszár, I.; Elias, P., Eds. János Bolyai Mathematical Society and North-Holland, 1977, Vol. 16, *Colloquia Mathematica Societatis János Bolyai*, pp. 493–519.
4. Cachin, C. Entropy measures and unconditional security in cryptography. PhD thesis, SWISS FEDERAL INSTITUTE of TECHNOLOGY ZURICH, 1997.
5. Renner, R.; Wolf, S. Advances in Cryptology—ASIACRYPT 2005: 11th International Conference on the Theory and Application of Cryptology and Information Security, Chennai, India, December 4–8, 2005. Proceedings; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Chapter Simple and Tight Bounds for Information Reconciliation and Privacy Amplification, pp. 199–216.
6. Hayashi, M. Exponential decreasing rate of leaked information in universal random privacy amplification. *Information Theory, IEEE Transactions on* **2011**, *57*, 3989–4001.
7. Cover, T.M.; Thomas, J.A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience, 2006.
8. Fehr, S.; Berens, S. On the conditional Rényi entropy. *Information Theory, IEEE Transactions on* **2014**, *60*, 6801–6810.

9. Iwamoto, M.; Shikata, J. Information theoretic security for encryption based on conditional Rényi entropies. In *Information Theoretic Security*; Springer, 2013; pp. 103–121.
10. Jizba, P.; Kleinert, H.; Shefaat, M. Rényi's information transfer between financial time series. *Physica A: Statistical Mechanics and its Applications* **2012**, *391*, 2971–2989.

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).