Theory of Approximation and Applications

Vol. 12, No.2, (2018), 17-28

A Preferred Definition of Conditional Rényi Entropy

Leila Golshani [a,*]

[a] *Department of Mathematics, Central Tehran Branch, Islamic Azad University, Tehran, Iran.*

Received 27 August 2018; accepted 22 December 2018

Abstract

In this paper, we focus on the definitions of conditional Rényi entropy, and select one of them on the basis of a relation between Rényi and Tsallis entropies, and show that the chain rule holds generally for the case of conditional Rényi entropy. Then, using this definition, we show some of the properties of conditional Rényi entropy. Finally, we show the relations among Rényi, Shannon and Tsallis entropies

Key words: Conditional entropy, Rényi entropy, Shannon entropy, Tsallis entropy

2010 AMS Mathematics Subject Classification : 62B10, 94A17.

* Corresponding author's E-mail: leila_ golshani@ yahoo.com(L.Golshani)

1 Introduction

The Rényi entropy is a generalization of Shannon entropy to a oneparameter family of entropies, by defining an entropy of order α. This was proposed by Rényi, in 1961 [16]. In 1988, Tsallis [21] too proposed a generalization of Shannon entropy (i.e., Tsallis entropy), by postulating a non-extensive entropy. This entropy is obtained in a form different from Rényi entropy. The measure for Tsallis entropy is non-logarithmic. After the introduction of Shannon entropy in 1948 [19], the conditional Shannon entropy was derived and its properties became known. Also, for Tsallis entropy, the conditional entropy was introduced [2,18] and its properties were shown [8]. What is mostly considered so far for the Rényi entropy is its underlying axioms, but no specific definition has been given for the conditional Rényi entropy.

Based on the definition of the conditional Shannon entropy, some authors such as Arimoto [1] and Cachin [4] gave definitions for the conditional Rényi entropy. In 2004, Jizba and Arimitsu [12] used the already introduced axioms for Shannon and Rényi entropies to introduce new axioms, on the basis of which one can get both Shannon and Rényi entropies, and also another definition for the conditional Rényi entropy is derived. Renner and Wolf [17], obtained the other definition by letting $\epsilon = 0$ in the conditional smooth Rényi entropy. In 2011, Hayashi [10] gave another definition, to derive an upper bound for the leaked information in universal privacy amplification. Recently, Fehr and Berens [7] reconsidered the definition of the conditional Rényi entropy which had been proposed by Arimoto [1] and showed that this particular notion satisfies several natural properties. In [11] and [20], the authors used some definitions of the conditional Rényi entropy, to find their properties and relations among them, but there is no general agreement on any specific definition. In this paper, we consider the definition of the conditional Rényi entropy given in [9]. To do this, we use the relation between Rényi and Tsallis entropies and show that the chain rule holds for the conditional Rényi entropy in general.

An operational use for the conditional Rényi entropy in cryptography was given in [4,11], and there were a number of applications in other fields,

such as quantum systems [22], biomedical engineering [14], economics [3], fields related to statistics [13], and other areas [6,15]. Thus, due to the large number of applications of conditional Rényi entropy, it is important to get a specific definition of it.

This paper is organized as follows. In section 2, we focus on the definitions of conditional Rényi entropy and we select one of them and then investigate some of its properties. In section 3, the relation among Shannon, Rényi and Tsallis entropies are given.

2 Conditinal Rényi entropy

The Shannon entropy of a probability distribution $P = (p_1, \ldots, p_n)$, or of a random variable X with probability distribution $P(X = x) = p(x)$, $x = 1, \ldots, n$, is defined as in [19], i.e.

$$H_1(X) \equiv H_1(P) = -\sum_x p(x) \ln p(x), \qquad \sum_x p(x) = 1 \tag{2.1}$$

and its Rényi entropy is given by [16]:

$$H_\alpha(X) \equiv H_\alpha(P) = \frac{1}{1-\alpha} \ln \left(\sum_x p^\alpha(x) \right), \quad \alpha > 0,\ \alpha \neq 1 \tag{2.2}$$

and its Tsallis entropy is given by [21], i.e.

$$S_\alpha(X) \equiv S_\alpha(P) = \frac{1}{1-\alpha} \left(\sum_x p^\alpha(x) - 1 \right), \quad \alpha > 0,\ \alpha \neq 1 \tag{2.3}$$

For Tsallis and Rényi entropies using (2.2) and (2.3) we have the following relation:

$$H_\alpha(X) = \frac{1}{1-\alpha} \ln(1 + (1-\alpha) S_\alpha(X)), \quad \alpha > 0,\ \alpha \neq 1 \tag{2.4}$$

Now, we review definitions of conditional Shannon entropy, conditional Tsallis entropy and conditional Rényi entropy.

The conditional Shannon entropy for random variable Y given X, with conditional probability distribution $P(Y = y|X = x) = p(y|x)$, $x = 1, \ldots, n$, $y = 1, \ldots, m$, is given by [5]:

$$H_1(Y|X) = \sum_x p(x) H_1(Y|X = x) = -\sum_{x,y} p(x)p(y|x) \ln p(y|x) \tag{2.5}$$

where $P(X = x) = p(x)$, $x = 1, \ldots, n$ is the probability distribution of X.

Abe [2] introduced the conditional Tsallis entropy in the following form:

$$\begin{aligned} S_\alpha(Y|X) &= \frac{S_\alpha(X,Y) - S_\alpha(X)}{1 + (1-\alpha)S_\alpha(X)} \\ &= \frac{1}{1-\alpha}\left[\frac{\sum_{x,y} p^\alpha(x,y)}{\sum_{x,y} p^\alpha(x)} - 1\right] \end{aligned} \tag{2.6}$$

where $p(x,y) = P(X = x, Y = y)$, $x = 1, \ldots, n$, $y = 1, \ldots, m$, is the probability distribution of random vector (X, Y), and $S_\alpha(X,Y)$ is the joint Tsallis entropy. Abe also obtained [2] pseudo-additivity property for Tsallis entropy, i.e.

$$S_\alpha(X,Y) = S_\alpha(X) + S_\alpha(Y|X) + (1-\alpha)S_\alpha(X)S_\alpha(Y|X) \tag{2.7}$$

and when X and Y are independent random variables, we have:

$$S_\alpha(X,Y) = S_\alpha(X) + S_\alpha(Y) + (1-\alpha)S_\alpha(X)S_\alpha(Y)$$

Several definitions for conditional Rényi entropy have been proposed, e.g., in [1,4,9,10,12,17]. In [7,11,20] some relations and properties of these definitions are discussed. These definitions are as follows:

(1) In [4],

$$\begin{aligned} H_\alpha^{(1)}(Y|X) &= \sum_x p(x) H_\alpha(Y|X = x) \\ &= \frac{1}{1-\alpha}\sum_x p(x) \ln \sum_y p^\alpha(y|x) \end{aligned} \tag{2.8}$$

(2) In [9] and [12],

$$H_\alpha^{(2)}(Y|X) = \frac{1}{1-\alpha} \ln \frac{\sum\limits_{x,y} p^\alpha(x,y)}{\sum\limits_{x} p^\alpha(x)} \tag{2.9}$$

(3) In [17],

$$H_\alpha^{(3)}(Y|X) = \frac{1}{1-\alpha} \ln \max_x \sum_y p^\alpha(y|x) \tag{2.10}$$

(4) In [1],

$$H_\alpha^{(4)}(Y|X) = \frac{\alpha}{1-\alpha} \ln \sum_x p(x) \left(\sum_y p^\alpha(y|x) \right)^{\frac{1}{\alpha}} \tag{2.11}$$

(5) In [10],

$$H_\alpha^{(5)}(Y|X) = \frac{1}{1-\alpha} \ln \left(\sum_x p(x) \sum_y p^\alpha(y|x) \right) \tag{2.12}$$

In [9], some reasons are given for taking (2.9) as the definition of the conditional Rényi entropy. In this paper, we consider some other arguments for using this definition.

Sanei Tabass et al. [18] consider the relation between Rényi and Tsallis entropies and the conditional Rényi entropy, proposed in [9], to obtain the conditional Tsallis entropy. This relation is the relation (2.4) that Abe [2] had obtained. Therefore, we see one reason for considering the relation (2.9) for the conditional Rényi entropy.

For conditional Shannon entropy, the chain rule and monotonicity are held [5].

According to chain rule we have:

$$H_1(Y|X) = H_1(X,Y) - H_1(X)$$

and according to monotonicity we have:

$$H_1(Y|X) \leq H_1(Y)$$

where $H_1(X,Y)$ is the joint Shannon entropy. Similar to the conditional Shannon entropy, it is natural to consider these properties for the conditional Rényi entropy. For the five definitions of conditional Rényi entropy already given, these properties are investigated. In the following, we review them, and for this purpose, the conditional Rényi entropy is denoted by $H_\alpha(Y|X)$.

(1) Chain rule: $H_\alpha(Y|X) = H_\alpha(X,Y) - H_\alpha(X)$ .This relation holds only for $H_\alpha^{(2)}(Y|X)$, [9,12].
(2) Monotonicity: $H_\alpha(Y|X) \leq H_\alpha(Y)$. This relation holds for $H_\alpha^{(4)}(Y|X)$, [1,7] and for $H_\alpha^{(5)}(Y|X)$ by theorem 4 in [11]. In general, it does not hold for $H_\alpha^{(2)}(Y|X)$, [12]. In [20, Theorem 7], it is pointed out that $H_\alpha^{(1)}(Y|X)$ and $H_\alpha^{(3)}(Y|X)$ do not satisfy this relation.

Now, in this section we show that the chain rule holds generally for the case of conditional Rényi entropy, and for this case the conditional Rényi entropy is denoted by $H_\alpha(Y|X)$.

Theorem 2.1 *For Rényi entropy, the chain rule holds, i.e.*

$$H_\alpha(Y|X) = H_\alpha(X,Y) - H_\alpha(X)$$

where $H_\alpha(X,Y)$ is the joint Rényi entropy.

Proof. By (2.4), we have for random vector (X,Y):

$$H_\alpha(X,Y) = \frac{1}{1-\alpha}\ln(1+(1-\alpha)S_\alpha(X,Y))$$

and by (2.7) we have:

$$\begin{aligned}
H_\alpha(X,Y) =& \frac{1}{1-\alpha}\ln\{1+(1-\alpha)[S_\alpha(X) \\
&+ S_\alpha(Y|X)+(1-\alpha)S_\alpha(X)S_\alpha(Y|X)]\} \\
=& \frac{1}{1-\alpha}\ln\{1+(1-\alpha)S_\alpha(X)+(1-\alpha)S_\alpha(Y|X) \\
&+ (1-\alpha)^2 S_\alpha(X)S_\alpha(Y|X)\} \\
=& \frac{1}{1-\alpha}\ln\{[1+(1-\alpha)S_\alpha(Y|X)][1+(1-\alpha)S_\alpha(X)]\} \\
=& \frac{1}{1-\alpha}\ln[1+(1-\alpha)S_\alpha(Y|X)] + \frac{1}{1-\alpha}\ln[1+(1-\alpha)S_\alpha(X)]
\end{aligned}$$

then by using the relation between Tsallis and Rényi entropies the result is obtained.

By this theorem, and without considering definitions of conditional Rényi entropy, the chain rule holds. Also by theorem 2.1, the conditional Rényi entropy proposed in [12] and [9] is obtained. So this is another argument for considering the relation (2.9) for the definition of the conditional Rényi entropy.

Now, we show some new properties of this definition of the conditional Rényi entropy.

Proposition 2.1 $H_\alpha(Y|X) \geq 0$, $\forall\ \alpha$.

Proof. In the relation (2.9), we can write $\sum_{x,y} p^\alpha(x,y)$ as $\sum_{x,y} p^\alpha(x,y) = \sum_x p^\alpha(x)\sum_y p^\alpha(y|x)$. So for $\alpha > 1$, $\sum_y p^\alpha(y|x) \leq \sum_y p(y|x) = 1$ and for $\alpha < 1$, $\sum_y p^\alpha(y|x) \geq \sum_y p(y|x) = 1$, and then the result is easily obtained.

Proposition 2.2 *$H_\alpha(Y|X) \leq H_\alpha(Y)$, $\forall\ \alpha$, where Y has a probability distribution, $P(Y=y)=\frac{1}{m}$, $y=1,\ldots,m$ the equality holds if and only if X and Y are independent.*

Proof. The complete proof of this relation is given in [12].

Theorem 2.2 *$H_\alpha(Y|X) \leq H_\alpha(Y)$, $\forall\ \alpha$, where X has a probability distribution $P(X = x) = \frac{1}{n}$, $x = 1, \ldots, n$ the equality holds if and only if X and Y are independent.*

Proof. This can be proved in a way similar to the case of the proposition 2.2.

Remark 2.1 *For two independent random variables X and Y we have: $p^\alpha(x, y) = p^\alpha(x)p^\alpha(y)$, thus $H_\alpha(Y|X) = H_\alpha(Y)$, $\forall\ \alpha$*

Remark 2.2 *The relation $H_\alpha(Y|X) \leq H_\alpha(Y)$, $\forall\ \alpha$, does not hold in general, unless X or Y has a uniform probability distribution.*

Theorem 2.3 ***(Chain rule for conditional entropy):*** *For random variables $X_1, X_2, \ldots, X_n$ and Y with the joint probability distribution $P(X_1 = x_1, \ldots, X_n = x_n, Y = y) = p(x_1, \ldots, x_n, y)$ we have:*

$$H_\alpha(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H_\alpha(X_i|X_1, \ldots, X_{i-1}, Y) \qquad \forall\ \alpha$$

Proof. By relation (2.9), we have:

$$H_\alpha(X_1, \ldots, X_n|Y) = \frac{1}{1-\alpha} \ln \frac{\sum\limits_{x_1,\ldots,x_n,y} p^\alpha(x_1, \ldots, x_n, y)}{\sum\limits_{y} p^\alpha(y)} \tag{2.13}$$

We can write:

$$\begin{aligned}
&\sum_{x_1,\ldots,x_n,y} p^\alpha(x_1, \ldots, x_n, y) \\
&= \sum_{x_1,y} p^\alpha(x_1, y) \frac{\sum\limits_{x_1,,x_2,y} p^\alpha(x_1, x_2, y)}{\sum\limits_{x_1,y} p^\alpha(x_1, y)} \cdots \frac{\sum\limits_{x_1,\ldots,x_n,y} p^\alpha(x_1, \ldots, x_n, y)}{\sum\limits_{x_1,\ldots,x_{n-1},y} p^\alpha(x_1, \ldots, x_{n-1}, y)}
\end{aligned}$$

Then, by inserting this equation into (2.13), we get:

$$\begin{aligned} H_\alpha(X_1,\ldots,X_n|Y) =& H_\alpha(X_1|Y) + H_\alpha(X_1|X_1,Y) + \cdots \\ &+ H_\alpha(X_n|X_1,\ldots,X_{n-1},Y) \\ =& \sum_{i=1}^{n} H_\alpha(X_i|X_1,\ldots,X_{i-1},Y) \end{aligned}$$

3 Relation among Rényi, Shannon and Tsallis entropies

In this section, we show the relation among Rényi, Shannon and Tsallis entropies. For this, we first consider the following property for the Rényi entropy.

Remark 3.1 *Rényi entropy $H_\alpha(X)$, for all X is a non-negative decreasing function of α, i.e. for $\alpha_1 < \alpha_2$, $H_{\alpha_2}(X) \leq H_{\alpha_1}(X)$ for all X, with the equality holding if and only if X is a uniform random variable.*

Using this remark, we have the following inequalities:

$$H_1(\cdot) \leq H_\alpha(\cdot), \qquad \alpha < 1 \tag{3.1}$$
$$H_\alpha(\cdot) \leq H_1(\cdot), \qquad \alpha > 1 \tag{3.2}$$

where H_1 is the Shannon entropy.

Now, we show the relation between Rényi and Tsallis entropies.

Lemma 3.1 ***(Fundamental inequality):*** *For any $a > 0$, $\ln a \leq a - 1$, with equality holding if and only if $a = 1$, [21].*

Proposition 3.1 *For Rényi and Tsallis entropies the following inequalities hold.*

$$H_\alpha(\cdot) \leq S_\alpha(\cdot), \qquad \alpha < 1 \tag{3.3}$$
$$H_\alpha(\cdot) \geq S_\alpha(\cdot), \qquad \alpha > 1 \tag{3.4}$$

Proof. $a = \sum_x p^\alpha(x)$ $(a > 0)$. Then by lemma 3.1, we have $\ln \sum_x p^\alpha(x) \leq \sum_x p^\alpha(x) - 1$. Let $\alpha < 1$, then by multiplying both sides of the relation through $\frac{1}{1-\alpha}$, we get $H_\alpha(\cdot) \leq S_\alpha(\cdot)$. In a similar way, the relation (3.4) is obtained for $\alpha > 1$.

Now, by remark 3.1 and proposition 3.1, the following inequalities are obtained among Rényi, Shannon and Tsallis entropies.

$$H_1(\cdot) \leq H_\alpha(\cdot) \leq S_\alpha(\cdot), \qquad \alpha < 1 \tag{3.5}$$
$$H_1(\cdot) \geq H_\alpha(\cdot) \geq S_\alpha(\cdot), \qquad \alpha > 1 \tag{3.6}$$

In the following, we illustrate these relations through an example.

Example 3.1 *Let X be a random variable with the probability distribution $P(X = 0) = 1 - P(X = 1) = \frac{1}{3}$. Then, the Shannon entropy is $H_1(X) = 0.63$, the Rényi entropy is, $H_\alpha(X) = \frac{1}{1-\alpha} \ln \left[\left(\frac{1}{3}\right)^\alpha + \left(\frac{2}{3}\right)^\alpha \right]$ and the Tsallis entropy is $S_\alpha(X) = \frac{1}{1-\alpha} \left[\left(\frac{1}{3}\right)^\alpha + \left(\frac{2}{3}\right)^\alpha - 1 \right]$. Let $\alpha = 2$, then $S_\alpha(X) = 0.44$ and $H_\alpha(X) = 0.59$, and for $\alpha = 0.5$, $S_\alpha(X) = 0.78$ and $H_\alpha(X) = 0.69$.*

4 Conclusions and Suggestions

In this paper, we have selected one of the five prevalent definitions of the conditional Rényi entropy and have demonstrated some of its properties. Also, we showed relation among Rényi, Tsallis and Shannon entropies. In the extension of this work, we can obtain other properties of this definition of the conditional Rényi entropy.

References

[1] S. Arimoto, Information measures and capacity of order α for discrete memoryless channels. Colloquia Mathematica Societatis Janos Bolyai, 16. Topics in Information Theory (1975) 41-52.

[2] S. Abe, Axioms and uniqueness theorem for Tsallis entropy, Phys. Lett. A. *271* (2000) 74-79.

[3] S. R. Bentes, R. Menezes, D.A. Mendes, Long memory and volatility clustering: is the empirical evidence consistent across stock markets?, Physica A *387* (2008) 3826-3830.

[4] C. Cachin, Entropy measures and unconditional security in cryptography. PhD thesis, Swiss Federal Institute of Technology Zurich, 1997.

[5] T. M. Cover, J. Thomas, The elements of information theory. John Wiley and Sons, 1991.

[6] A. Dukkipati, S. Bhatnagar, M.N. Murty, Gelfand-Yaglom-Perez theorem for generalized relative entropy functionals, Information Sciences *177* (2007) 5707-5714.

[7] S. Fehr, S. Berens, On the Conditional Rényi Entropy. IEEE Transactions on Information Theory, *60* (2014) 6801-6810.

[8] S. Furuichi, Information theoretical properties of Tsallis entropies, Journal of Mathematical Physics. *47* (2006).

[9] L. Golshani, E. Pasha, G. Yari, Some properties of Rényi entropy and Rényi entropy rate, Information sciences. *179* (2009) 2426-2433.

[10] M. Hayashi, Exponential decreasing rate of leaked information in universal random privacy amplification. IEEE Transactions on Information Theory *57* (2011) 3989-4001.

[11] M. Iwamoto, J. Shikata: Information Theoretic Security for Encryption Based on Conditional Rényi Entropies. IACR Cryptology ePrint Archive 2013: *440* (2013).

[12] P. Jizba, T. Arimitsu, The world according to Rényi: thermodynamics of multifractal systems. J. Ann.Phys. *312* (2004) 17-59.

[13] F. Kanaya, T.S. Han, The asymptotics of posterior entropy and error probability for Bayesian estimation, IEEE Transactions on Information Theory *41* (1995) 1988-1992.

[14] D.E. Lake, Rényi entropy measures of heart rate gaussianity, IEEE Transactions on Biomedical Engineering *53* (2006) 21-27.

[15] J. B. Paris, S. R. Rad, Inference processes for quantified predicate knowledge, Logic, Language, Information and Computation. *5110* (2008) 249-259.

[16] A. Rényi, On measures of entropy and information, in proc .4th Berkeley symp.mathematical statistics probability. Berkeley, CA: univ .calif .press. *1* (1961), 547-561.

[17] R. Renner, S. Wolf, Simple and tight bounds for information reconciliation and privacy amplification. Advances in Cryptology- ASIACRYPT 2005, LNCS4515, Springer-Verlag (2005) 199-216.

[18] M. Sanei Tabass, M. G. Mohtashami Borzadaran, M. Amini, (2013), Conditional Tsallis Entropy. Cybernetics and Information Technologies, Bulgarian Academy of Sciences, *13*, 37-42.

[19] C. E. Shannon, A mathematical theory of communication. Bell Syst .Techn. J. *27* (1948), 379-423, 623-656.

[20] A. Teixeira, A. Matos, L. Antunes, Conditional Rényi entropies. IEEE Trans. Information Theory *58* (2012) 4273-4277.

[21] C. Tsallis, Possible generalization of Bolzmann-Gibbs statistics, J. Stat. Phys, *52* (1988) 479-487.

[22] K. G. H. Vollbrecht, M. M. Wolf, Conditional entropies and their relation to entanglement criteria, Journal of Mathematical Physics *43* (2002) 4299-4306.