BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **13**, No 2

Sofia • 2013 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2013-0012

Conditional Tsallis Entropy

Sanei Tabass Manije, Mohtashami Borzadaran Gholamreza, Amini Mohammad

Department of Statistics, Faculty of Mathematical Sciences, The Ferdowsi University of Mashhad, Mashhad-IRAN
Emails: manije_sanei@yahoo.com gmb1334@yahoo.com mamini48@yahoo.com

Abstract: *In this paper, the conditional Tsallis entropy is defined on the basis of the conditional Renyi entropy. Regarding the fact that Renyi entropy is the monotonically increasing function of Tsallis entropy, a relationship has also been presented between the joint Tsallis entropy and conditional Tsallis entropy.*

Keywords: *Shannon entropy, Conditional Shannon entropy, Conditional Tsallis entropy, Conditional Renyi entropy, Chain rule.*

1. Introduction

Experiences of Shannon during the Second World War led to the revolution in communication theory and he then introduced Shannon entropy [13] in 1948. Then, scientists (e.g., Renyi [12], Tsallis [14]) generalized this idea and obtained different entropic forms which, in special cases, include Shannon entropy. Renyi [12] in 1961 generalized Shannon entropy by defining an entropy which is called the Renyi entropy. Harvda and Charvat [9] in 1967 defined an extension of entropy and in 1988 Tsallis [14] too proposed the generalization of the entropy by postulating a non-extensive entropy, (i.e., Tsallis entropy), which covers Shannon entropy in particular cases. This measure is non-logarithmic. These forms of entropy are obtained through the joint generalization of the averaging procedures and the concept of information gain. Renyi entropy as well as Shannon entropy preserves the same definition of information gain, however, it makes use of exponential average where as Shannon entropy makes use of common average procedure. On the other hand, Tsallis entropy is not extensive [15] but generalizes the concept of the information gain and is obtained by the linear averaging procedure. In fact, Renyi entropy is a monotonically increasing function of Tsallis entropy. We know that conditional Shannon entropy has been introduced and for random variables and that there is a relation between conditional and joint Shannon entropies for random variables, indeed for which the chain rule holds. Cachin [4] has given a definition of conditional Renyi entropy that chain rule does not hold. Then Jizba and Arimitsu [11] introduced some axioms for Renyi entropy and Golshani et al. [8], preserved different form of conditional Renyi entropy and showed that chain rule holds for that by using their axioms.

In this paper, we have introduced a form of conditional Tsallis entropy on the basis of the conditional Renyi entropy in view of Cachin [4] and then shown that this definition is not suitable. So, we used the axioms introduced by Jizba and Arimitsu [11] for Renyi entropy and monotone relation between Renyi and Tsallis entropy and gave a different form of conditional Tsallis entropy. in the end, given the fact that Tsallis entropy is not additive, we have proved a rule similar to chain rule for Tsallis entropy.

2. Tsallis entropy

Renyi entropy and Tsallis entropy play a central role in different sciences as extensive forms of the Shannon entropy. For example see [1-3, 7, 10-15]. In the following, a definition for Tsallis conditional entropy of random variables having discrete distribution, as well as its properties are presented. The Shannon entropy of a probability distribution $P=(p_1, p_2, \ldots, p_n)$ or of a random variable X, with probability distribution $P(X=x_i)=p_i$, $i=1, 2, \ldots, n$, is defined as

(2.1) $$H_1(X) \equiv H_1(P) = -\sum_i p_i \log p_i = \sum_i p_i(-\log p_i)$$

where relation (2.1) is the arithmetic mean of the information $-\log p_i$. Also, its Renyi entropy is

(2.2) $$H_\alpha(X) \equiv H_\alpha(P) = \frac{1}{1-\alpha} \log \sum_i p_i^\alpha, \quad \alpha > 0,\ \alpha \neq 1.$$

Renyi entropy becomes the Shannon entropy as $\alpha \to 1$. This relation for a random vector (X, Y) with probability distribution $P(X=x_i, Y=y_j)=p_{ij}$, $i=1, 2, \ldots, n$, and $j=1, 2, \ldots, m$, can be written as

(2.3) $$H_\alpha(X,Y) \equiv H_\alpha(P) = \frac{1}{1-\alpha} \log \sum_{i,j} p_{ij}^\alpha, \quad \alpha > 0,\ \alpha \neq 1.$$

Tsallis entropy (which is simply related to Renyi entropy) is another generalized entropy measure based on the generalization of the information gain. This measure is defined as follows:

Definition 2.1. Tsallis entropy of a probability distribution $P=(p_1, \ldots, p_n)$ or of a random variable X, with probability distribution $P(X=x_i)=p_i$, $i=1, 2, \ldots, n$, is defined as

$$(2.4)\qquad S_\alpha(X) \equiv S_\alpha(P) = \frac{1}{\alpha-1}[1-\sum_i p_i^\alpha],\quad \alpha>0, \alpha \neq 1.$$

If we extend this definition to the case of random vector (X, Y) with probability distribution $P(X = x_i, Y = y_j) = p_{ij}$, $i = 1, ..., n$, and $j = 1, ..., m$, then the joint Tsallis entropy is

$$(2.5)\qquad S_\alpha(X,Y) \equiv S_\alpha(P) = \frac{1}{\alpha-1}[1-\sum_{i,j} p_{ij}^\alpha],\quad \alpha > 0,\ \alpha \neq 1.$$

In fact, Renyi entropy can be shown [4] to be a monotonically increasing function of Tsallis entropy and the two are related via the following relation:

$$(2.6)\qquad H_\alpha(X) = \frac{1}{1-\alpha}\log[1+(1-\alpha)S_\alpha(X)].$$

Now we seek a suitable definition for conditional Tsallis entropy by relation (2.6). First of all, an overview of conditional Shannon entropy is provided below. For random variable Y, given X, with conditional probability distribution $P(Y= y_j|X = x_i) = p_{j|i}$, we have:

$$(2.7)\qquad H_1(Y\mid X) = \sum_i p_i H(Y \mid X = x_i) = -\sum_i p_i p_{j|i}\log p_{j|i}.$$

On the basis of the conditional Shannon entropy (2.7), we propose the following definition for conditional Tsallis entropy:

$$(2.8)\qquad S_\alpha(Y\mid X) = \sum_i p_i \frac{1}{\alpha-1}[1-\sum_j p_{j|i}^\alpha] = \frac{1}{\alpha-1}\sum_i p_i[1-\sum_j p_{j|i}^\alpha].$$

Although Tsallis entropy, just like Shannon entropy, employs the same averaging procedure called linear averaging, it seems that the relation (2.8) must be suitable, which is disproved in the next section of the current study. Following Jizba and Arimitsu [11], Golshani et al. [8] presented the following definition for conditional Renyi entropy. For probability distributions $P = (p_1, ..., p_n)$, $Q = (q_1, ..., q_m)$ and the joint probability distribution denoted by p_{ij}, the conditional Renyi entropy of random variable Y, given X, is defined as

$$(2.9)\qquad H_\alpha(Y\mid X) = \frac{1}{\alpha-1}\log\frac{\sum_{i,j} p_{ij}^\alpha}{\sum_i p_i^\alpha},$$

and for this equation $H_\alpha(Y|X) = H_\alpha(X, Y) - H_\alpha(X)$.

3. Conditional Tsallis entropy

Since Renyi and Tsallis entropy are related as mentioned in (2.6), now we present another definition for conditional Tsallis entropy.

Definition 3.1. The conditional Tsallis entropy of random variable Y given X is defined as

$$(2.10)\qquad S_\alpha(Y\mid X) = \frac{1}{\alpha-1}\left[1-\frac{\sum_{i,j} p_{ij}^\alpha}{\sum_i p_i^\alpha}\right].$$

Using this definition, we can show that the conditional Shannon entropy as $\alpha \to 1$, also via (2.6) and (2.10) we can obtain:

$$S_\alpha(X,Y) = \frac{1}{\alpha - 1}\left[1 - \sum_{i,j} p_{ij}^\alpha\right] = \frac{1}{\alpha-1}[1 - \exp((1-\alpha)H_\alpha(X,Y))] =$$

$$(2.11) \qquad = \frac{1}{\alpha-1}[1 - \exp((1-\alpha)H_\alpha(X) + H_\alpha(Y \mid X))] =$$

$$= \frac{1}{\alpha-1}[1 - [1 + (1-\alpha)S_\alpha(X)][1 + (1-\alpha)S_\alpha(Y \mid X)]] =$$

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

By considering the relation (2.10) we see that this conditional Tsallis entropy verifies in (2.11), and we can show through one example that the equality holds. It does not hold for the relation (2.8) for conditional Tsallis entropy, so we can reach the conclusion that relation (2.10) could better define conditional Tsallis entropy rather than the relation (2.8) does.

Example 3.2. Let X and Y be two random variables with the joint distribution $P(0,\ 0) = 0$, $P(0,\ 1) = P(1,\ 0) = P(1,\ 1) = 1/3$, the conditional Tsallis entropy is obtained from equation (2.8) and it is given by $S_\alpha(Y \mid X) = \frac{1}{\alpha-1}[1 - 1/3 + 4/3(1/2^\alpha)]$ and Tsallis entropy for random variable X is $S_\alpha(X) = \frac{1}{\alpha-1}\left[1 - \frac{1+2^\alpha}{3^\alpha}\right]$. and the joint Tsallis entropy, using (2.5), is $S_\alpha(X,Y) = \frac{1}{\alpha-1}[1 - 3(1/3)^\alpha]$, thus relation (2.8) is not suitable.

Now we express some properties of the conditional Tsallis entropy.

Lemma 3.3. When Y has a probability distribution $P(Y = y_j) = 1/m$, $j = 1, 2, \dots, m$, we have $S_\alpha(Y \mid X) \le S_\alpha(Y)$, $\forall \alpha$, and the equality holds if X and Y are independent.

Proof: We have $p_{ij}^\alpha = p_j^\alpha p_{i|j}^\alpha = \left(\frac{1}{m}\right)^\alpha p_{i|j}^\alpha$, on the other hand $p_i^\alpha \le p_{i|j}^\alpha$, therefore:

$$\begin{cases} \sum_{i,j} p_{i|j}^\alpha \ge m \sum_i p_i^\alpha, \\ 1 - \left(\frac{1}{m}\right)^\alpha \frac{\sum_{i,j} p_{i|j}^\alpha}{\sum_i p_i^\alpha} \le 1 - \left(\frac{1}{m}\right)^\alpha m. \end{cases}$$

So we have $S_\alpha(Y \mid X) \le S_\alpha(Y)$.

Remark 3.4. For two independent random variables, we have $p_{ij} = p_i p_j$, therefore,

$$S_\alpha(Y \mid X)=\frac{1}{\alpha-1}\left[1-\frac{\sum_{i,j} p_i^\alpha p_j^\alpha}{\sum_i p_i^\alpha}\right]=\frac{1}{\alpha-1}\left[1-\sum_j p_j^\alpha\right]=S_\alpha(Y)\ \forall\alpha$$

when the random variable X takes values 1, 2, ... we extend Definition 2.1. So

(2.12) $$S_\alpha(X)=\frac{1}{\alpha-1}[1-\sum_{i=1}^{\infty} p_i^\alpha],\ \ \alpha>0,\ \alpha\neq 1,$$

(2.13) $$S_1(X)=-\sum_{i=1}^{\infty} p_i \log p_i.$$

If the series on the right-hand sides of (2.12) and (2.13) converge, these entropies exist. If $\alpha > 1$, summation (2.12) always converges, but summation (2.13) does not. If all entropies converge for $0 < \alpha < 1$, then all of the definitions and propositions that hold for Tsallis entropy of random variables with finite values, hold for infinite values as well. This issue is absolutely the same for Renyi entropy. For Shannon entropy the chain rule holds [5], and for the Renyi entropy, Golshani et al. [8] showed that the chain rule holds, which is expressed as follows:

Theorem 3.5. Let $(X_1, ..., X_n)$ be a random vector with probability distribution $P(i_1, ..., i_n)$ and $H_\alpha(X_1, ..., X_n)$ be the Renyi entropy, then

(2.14) $$H_\alpha(X_1,...,X_n)=\sum_{i=1}^{n} H_\alpha(X_i \mid X_1,...,X_{i-1}).$$

Now, considering the point that "Tsallis entropy is not extensive", we may obtain a relation between conditional Tsallis entropy and the joint Tsallis entropy, by the following theorem.

Theorem 3.6. Under assumptions of Theorem 3.5 and $S_\alpha(X_1, ..., X_n)$ be the Tsallis entropy. Then

(2.15) $$S_\alpha(X_1,...,X_n)=\frac{1}{\alpha-1}[1-\prod_{i=1}^{n}(1+(1-\alpha))S_\alpha(X_i \mid X_1,...,X_{i-1})].$$

Proof: For the random vector $(X_1, ..., X_n)$, we have:

$$S_\alpha(X_1,...,X_n)=\frac{1}{\alpha-1}\left[1-\sum_{i_1,...,i_n}^{n} p_{i_1,...,i_n}^\alpha\right].$$

Now via Theorem 3.5 and relation (2.6) we obtain

$$\frac{1}{1-\alpha}\log[(1+(1-\alpha)S_\alpha(X_1,...,X_n)]=\frac{1}{1-\alpha}\log[(1+(1-\alpha)S_\alpha(X_1)]+$$

$$+\frac{1}{1-\alpha}\log[(1+(1-\alpha)S_\alpha(X_2 \mid X_1)]+$$

$$\vdots$$

$$+\frac{1}{1-\alpha}\log[(1+(1-\alpha)S_\alpha(X_n \mid X_1,...,X_{n-1})].$$

hence

$$\log[(1+(1-\alpha)S_\alpha(X_1,...,X_n)] = \sum_{i=1}^{n}\log[(1+(1-\alpha)S_\alpha(X_i \mid X_1,...,X_{i-1})] =$$

$$=\log\prod_{i=1}^{n}[(1+(1-\alpha)S_\alpha(X_i \mid X_1,...,X_{i-1})].$$

And so, $S_\alpha(X_1,...,X_n) = \dfrac{1}{\alpha-1}[1-\prod_{i=1}^{n}(1+(1-\alpha)S_\alpha(X_i \mid X_1,...,X_{i-1})].$

4. Conclusion

Following J i z b a and A r i m i t s u [11], G o l s h a n i et al. [8] provided a definition for conditional Renyi entropy. Given that Renyi entropy is a monotonically increasing function of Tsallis entropy, the conditional Tsallis entropy was introduced in this paper on the basis of the conditional Renyi entropy. Also, we have obtained a relation between the joint Tsallis entropy and conditional Tsallis entropy.

Acknowledgment: The support of Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged.

R e f e r e n c e s

1. A n d a i, A. On the Geometry of Generalized Gaussian Distribution. – Journal of Multivariate Analysis, Vol. **100**, 2009, 777-793.
2. B a g s i, G. B., U. T i r n a k l i. On the Way Towards a Generalized Maximization Procedure. – In: Physics Letters A, Vol. **373**, 2009, 3230-3234.
3. B e r c h e r, J. F. Tsallis Distribution as a Standard Maximum Entropy Solution with Tail Constraint. – In: Physics Letters A, Vol. **372**, 2008, 5657-5659.
4. C a c h i n, C. Entropy Measures and Unconditional Security in Cryptography. PhD Thesis, Swiss Federal Institute of Thechnology, Zurich, 1997.
5. C o v e r, T. M., J. A. T h o m a s. Elements of Information Theory. Second Edition. Wiley Intern. Science, 2006.
6. C s i s z a r, I. Generelized Cut-Off Rates and Renyi Information Measures. – IEEE Trans. Inf. Theory, Vol. **41**, 1995, 26-34.
7. G r e g o r i o, A. D., S. M. L a c u s. On Renyi Information for Ergodic Diffusion Processes. – Information Sciences, 2009, 279-291.
8. G o l s h a n i, L., E. P a s h a, G. Y a r i. Some Properties of Renyi Entropy and Renyi Entropy 263 Rate. – Information Sciences, Vol. **179**, 2009, 2426-2433.
9. H a r v d a, J., F. C h a r v a t. Quantification Method of Classification Processes: Concept of Structural α-Entropy. – Kybernetika, Vol. **3**, 1967, 30-35.
10. J e n s s e n, R., T. E l t o f t. A New Information Theoretic Analysis of Sum of Squared Error Kernel Clustering. – Neurocomputing, Vol. **72**, 2008, 23-31.
11. J i z b a, P., T. A r i m i t s u. The World According to Renyi Thermodynamics of Multifractal Systems. – Annals of Physics, Vol. **312**, 2004, 17-59.
12. R e n y i, A. On Measures of Entropy and Information. – In: Proc. of Berekeley Symposium, Statist. Probability, 1, 1961, 547-561.
13. S h a n n o n, C. E. A Mathematical Theory of Communication Bellsystem. – Tech., Vol. **27**, 1948, 379-423.
14. T s a l l i s, C. Possible Generalizations of Boltzmann-Gibbs Statistics. – Journal of Statistical Physics, Vol. **52**, 1988, 479-487.
15. T s a l l i s, C. Entropic Nonextensivity: A Possible Measure of Complexity. – Chaos, Solitons and Fractals, Vol. **13**, 2002, 371-391.