Renyi entropy in continuous case is not the limit of discrete case

M. Sanei Tabass, G. R. Mohtashami Borzadaran, M. Amini

(Communicated by K. Şanlı KULA)

Abstract

Shannon entropy is the creation of Shannon (1948) based on the experiences in Bell System Company during and after the Second World War. Then, Renyi (1961) generalized it for one parameter families of entropies. This entropy for discrete random variables is non-negative but it can be negative in continuous case.

In this paper, we show that Renyi entropy for continuous random variables is not equal to the limit of it for discrete random variables. Also, some notes are derived in view of the variate versions of entropy criteria.

Keywords: Shannon entropy, Renyi entropy, Continuous random variable, Riemann integrable, Information Theory.

AMS Subject Classification (2010): Primary: 62B10 ; Secondary: 62E10.

1. Introduction and Preliminaries

Entropy describes the amount of uncertainty associated with a random variable. Information theory has mathematical origin in entropy notion that is related to Thermodynamic and Statistical Mechanics. Hartley (1928) presented definitions of information theory in engineering and for this reason, information theory has been taken as subdivision of the communication theory. Then, Shannon (1948) presented mathematical definition of entropy and succeed to determine the capacity of channel for transmission of numerical information using information sources. To improve his attempt on information theory, other scientists arrived to the Shannon entropy via considering different properties. For example, we can refer to Fadeev (1956), Tverberg (1958), Lee (1964), Didrrich (1975), Darroczy and Maksa (1979) and Aczel (1970).

After 1948, variate extensions of the Shannon entropy has been introduced such as Renyi entropy (1961), Harvda Charvat (1967) and Tsallis entropy (1988), which included Shannon entropy as a special case. Two other measures are exponential and generalized exponential entropies which were introduced by Koski and Persson (1992) and Campbell (1966).

Renyi and Tsallis entropies converge to the Shannon entropy. The generalized exponential entropy can be related to a generalized version of the Renyi entropy, generalized Renyi entropy and generalized exponential entropy converge to the Shannon entropy .

Renyi information divergence, Renyi mutual information are two other information measures that converge to the Kullback-Leibler information and mutual information respectively. We summarize these information measures in Table 1.

In this paper, we show that Renyi entropy for continuous random variables is not equal to the limit of it for discrete random variables. These results are hold for Tsallis entropy (by use of monotone relation that exists between these two measures of entropy) and generalized Renyi entropy. However, the Renyi mutual information between two

random variables is the limit of the mutual information between their quantized versions.

Table 1. Discrete and Continuous entropy and Generalizations.

Information measure	discrete	Continuous
Shannon entropy	$H(X) = -\sum_x p_x \log p_x$	$H(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx$
Kullback-Leibler	$D(P\|Q) = \sum_{x,y} P_{(x,y)} \log(\frac{P_{(x,y)}}{Q_{(x,y)}})$	$D(f\|g) = \int_{x,y} f(x,y) \log(\frac{f(x,y)}{g(x,y)}) dxdy$
Mutual information	$I(X,Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$	$I(X,Y) = \int_{x,y} f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy$
Renyi entropy	$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_x p_x^\alpha$ $\alpha > 0, \alpha \neq 1$	$H_\alpha(X) = \frac{1}{1-\alpha} \log \int_x f^\alpha(x) dx$ $\alpha > 0, \alpha \neq 1$
Renyi information Divergence	$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \log \sum_{x,y}$ $(\frac{P(x,y)}{Q(x,y)})^{\alpha-1} P(x,y)$	$D_\alpha(f\|g) = \frac{1}{\alpha-1} \log \int_{x,y}$ $(\frac{f(x,y)}{g(x,y)})^{\alpha-1} f(x,y) dxdy$
Renyi Mutual Information	$MD_\alpha(X,Y) = \frac{1}{\alpha-1} \sum_{x,y}$ $(\frac{P(x,y)}{P(x)P(y)})^{\alpha-1} P(x,y)$	$MD_\alpha(X,Y) = \frac{1}{\alpha-1} \int_{x,y}$ $(\frac{f(x,y)}{f(x)f(y)})^{\alpha-1} f(x,y) dxdy$
Tsallis entropy	$S_\alpha(X) = \frac{1}{\alpha-1} \sum_x p_x^\alpha - 1$ $\alpha > 0, \alpha \neq 1$	$S_\alpha(X) = \frac{1}{\alpha-1} \int_x f^\alpha(x) dx - 1$ $\alpha > 0, \alpha \neq 1$
Generalized Renyi entropy	$H_{\alpha,\beta}^{GR}(X) = \begin{cases} \frac{1}{\beta-\alpha} \ln \frac{\sum_x p_x^\alpha}{\sum_x p_x^\beta}, \alpha \neq \beta \\ \frac{\sum_x p_x^\alpha \ln p(x)}{\sum_x p_x^\alpha}, \alpha = \beta \end{cases}$	$H_{\alpha,\beta}^{GR}(X) = \begin{cases} \frac{1}{\beta-\alpha} \ln \frac{\int_x f^\alpha(x) dx}{\int_x f^\beta(x) dx}, \alpha \neq \beta \\ \frac{\int_x f^\alpha(x) \ln f(x) dx}{\int_x f^\alpha(x) dx}, \alpha = \beta \end{cases}$
Generalized exponential	$H_{\alpha,\beta}^{GE}(X) = exp(H_{\alpha,\beta}^{GR}(X))$	$H_{\alpha,\beta}^{GR}(X) = exp(H_{\alpha,\beta}^{GR}(X))$

2. Main Results

A wide variety of competing measures of entropy have emerged in other papers, a number of these measures that can be considered as generalizations of the Shannon entropy in continuous and discrete cases, are shown in Table 1. We suppose X as a random variable with density function $f(x)$ and pmf $P(x)$ for continuous and discrete random variables respectively.

Shannon entropy is always non-negative for discrete random variables, but it can be negative for continuous case. It is because of the Shannon entropy for continuous random variables is not equal to limit of the Shannon entropy for discrete random variables, this limit is infinite.

In what follows, firstly we express a theorem in view of comparing the Shannon entropy in continuous and discrete cases, (see Cover and Thomas 2006). Then, we extend it to Renyi and Tsallis cases.

Theorem 2.1. *Let X be a continuous random variable with density function $f(x)$ which is Riemann integrable, suppose that we divide the range of X into bins of the length Δ, consider the quantized random variable X^Δ which is defined by*

$$X^\Delta = x_i \quad if \quad i\Delta \leq X < (i+1)\Delta,$$

then,

$$\lim_{\Delta \to 0} \{H(X^\Delta) + \log\Delta\} = H(X). \tag{2.1}$$

Proof: See page 248 of Cover and Thomas (2006) for more details.

From the definition of mutual information it is clear that,

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

If $(X,Y) \sim p(x,y)$, the conditional Shannon entropy, is defined as

$$H(X|Y) = \sum_y p_y H(X|Y=y) = -\sum_y \sum_x p(x,y) \log p(x|y).$$

In continuous case the summation is replaced by integrals.

Remark 2.1. (Cover and Thomas 2006) The mutual information of two random variables is the limit of the mutual information between their quantized versions as follow:

$$\begin{aligned} I(X^\Delta, Y^\Delta) &= H(X^\Delta) + H(X^\Delta | Y^\Delta) \\ &\approx H(X) - log(\Delta) + H(X|Y) + log(\Delta) = I(X, Y). \end{aligned} \tag{2.2}$$

The Renyi entropy for discrete random variables is always non-negative, but in continuous case it can be negative. We express and prove the following theorem, in this direction:

Theorem 2.2. *Consider a random variable X with density $f(x)$ illustrated in Figure 1 and $f(x)$ is Riemann integrable, then we have:*

$$\lim_{\Delta \to 0} \{H_\alpha(X^\Delta) + log\Delta\} = H_\alpha(X). \tag{2.3}$$
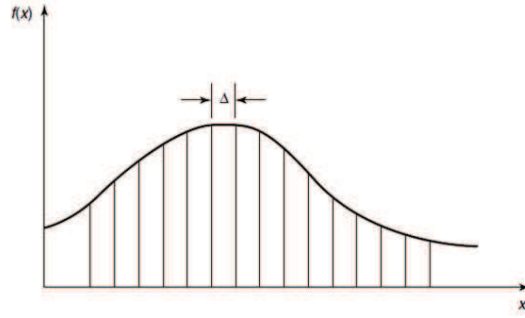

Figure 1. *Quantization of a continuous random variable (Cover and Thomas (2006))*

Proof: Suppose that we divide the range of X into bins of length Δ. Now consider the quantized random variable X^Δ which is defined by,

$$X^\Delta = x_i \quad if \quad i\Delta \le X < (i+1)\Delta.$$

The density function is continuous within the bins. By the mean value theorem, there exists a value x_i within each bin such that,

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx.$$

Then the probability that $X^\Delta = x_i$ is,

$$p_i = \int_{i\Delta}^{(i+1)\Delta} f(x)dx = f(x_i)\Delta.$$

Now we calculate The Renyi entropy for discrete random variable X^Δ:

$$\begin{aligned} H_\alpha(X^\Delta) &= \frac{1}{1-\alpha} log \sum_i p_i^\alpha = \frac{1}{1-\alpha} log \sum_i (f(x_i)\Delta)^\alpha \\ &= \frac{1}{1-\alpha} log \sum_i (f(x_i))^\alpha + \frac{\alpha}{1-\alpha} log\Delta \\ &= \frac{1}{1-\alpha} log \sum_i (f(x_i))^\alpha + \frac{\alpha}{1-\alpha} log\Delta + \frac{1}{1-\alpha} log\Delta - \frac{1}{1-\alpha} log\Delta \\ &= \frac{1}{1-\alpha} log \sum_i (f(x_i))^\alpha \Delta + \frac{\alpha-1}{1-\alpha} log\Delta \\ &= \frac{1}{1-\alpha} log \sum_i (f(x_i))^\alpha \Delta - log\Delta. \end{aligned} \tag{2.4}$$

Since function $(f(x))^\alpha$ is Riemann integrable, the first term in (2.4) approaches to:

$$\frac{1}{1-\alpha} log \int (f(x))^\alpha dx, \quad as\ \Delta \to 0.$$

In other word,

$$\lim_{\Delta \to 0} \frac{1}{1-\alpha} log \sum_i (f(x_i))^\alpha \Delta \to \frac{1}{1-\alpha} log \int (f(x))^\alpha dx,$$

Hence:

$$\lim_{\Delta \to 0} \{H_\alpha(X^\Delta) + log\Delta\} = H_\alpha(X).$$

This conclusion implies Equation (2.1) as $\alpha \to 1$.

Corollary 2.1. *Theorem 2.2 holds for Tsallis entropy also as follows:*

$$\lim_{\Delta \to 0} S_\alpha(X^\Delta) = S_\alpha(X)\Delta^{\alpha-1} + \frac{1}{\alpha-1},$$

We can prove this corollary according to the fact, "Renyi entropy is a monotonically increasing function of Tsallis entropy" (see for more details Renyi, 1961) and related by following relation

$$H_\alpha(X) = \frac{1}{1-\alpha} log[1 + (1-\alpha)S_\alpha(X)]. \tag{2.5}$$

Remark 2.2. Theorem 2.2 holds for both Generalized Renyi and Generalized exponential entropies and we can say that both of them in continuous cases are not the limit of the quantized versions. since,

$$
\begin{aligned}
H_{\alpha,\beta}^{GR}(X^\Delta) &= \frac{1}{\beta-\alpha} log \frac{\sum_i p_i^\alpha}{\sum_i p_i^\beta} \\
&= \frac{1}{\beta-\alpha} log \frac{\sum (f(x_i)\Delta)^\alpha}{\sum (f(x_i)\Delta)^\beta} \\
&= \frac{1}{\beta-\alpha} log \frac{\sum (f(x_i))^\alpha}{\sum (f(x_i))^\beta} + \frac{1}{\beta-\alpha} log\Delta^{\alpha-\beta} \\
&= \frac{1}{\beta-\alpha} log \frac{\sum (f(x_i))^\alpha \Delta}{\sum (f(x_i))^\beta \Delta} - log\Delta,
\end{aligned}
\tag{2.6}
$$

Function $(f(x))^\alpha$ is Riemann integrable, so if $\Delta \to 0$ relation (2.6) approaches to:

$$\frac{1}{\beta-\alpha} log \frac{\int_{-\infty}^{\infty} f^\alpha(x)dx}{\int_{-\infty}^{\infty} f^\beta(x)dx} - log\Delta,$$

And we have:

$$\lim_{\Delta \to 0} \{H_{\alpha,\beta}^{GR}(X^\Delta) + log\Delta\} = H_{\alpha,\beta}^{GR}(X). \tag{2.7}$$

Generalized exponential is increasing function of generalized Renyi entropy and Equation (2.7) holds for it.

Theorem 2.3. *The Renyi mutual information between two random variables is the limit of the mutual information between their quantized versions.*
We show this topic for Renyi information divergence, then conclude the Renyi mutual information between two random variables is the limit of the mutual information between their quantized versions via the following relations:

$$
\begin{aligned}
D_\alpha(X^\Delta, Y^\Delta) &= \frac{1}{\alpha-1} log \sum_i (\frac{p_i}{q_i})^{\alpha-1} p_i \\
&= \frac{1}{\alpha-1} log \sum_i (\frac{f(x_i)\Delta}{g(x_i)\Delta})^{\alpha-1} f(x_i)\Delta \\
&= \frac{1}{\alpha-1} log \sum_i (\frac{f(x_i)}{g(x_i)})^{\alpha-1} f(x_i)\Delta.
\end{aligned}
\tag{2.8}
$$

Since $(\frac{f(x)}{g(x)})^{\alpha-1}f(x)$ is Riemann integrable, we obtain:

$$lim_{\Delta \to 0} \frac{1}{\alpha-1} log \sum_i (\frac{f(x_i)}{g(x_i)})^{\alpha-1} f(x_i)\Delta = \frac{1}{\alpha-1} log \int (\frac{f(x)}{g(x)})^{\alpha-1} f(x) dx.$$

Corollary 2.2. *Now by applying relation (2.8) and $MD_\alpha(X, Y)$ as have shown in Table 1, we come to this conclusion that Renyi mutual information between two random variables is the limit of the Renyi mutual information between their quantized versions.*

3. Conclusion

In this article, we have shown that Renyi entropy for continuous random variables is not limit of it for discrete case. Then we have proved this result for other generalized entropies, including Tsallis and generalized Renyi entropy. Also, we have found that Renyi divergence, Renyi mutual information in continuous cases are limit of their discrete cases.

References

[1] Aczel, J. Forte, B. and Ng, C. T., Why Shannon and Hartley entropies are "natural". *Adv. Appl. Probab.* (1970), 6, 131-146.

[2] Campbell, L.L., Exponential entropy as a measure of extent of a distribution. *Zeitschr. fur Wahrsch. und verw. Geb.* (1966), 5, 217-255.

[3] Cover, T. M. and Thomas, J. A., Elements of Information Theory. Second Edition. *Wiley Interscience.* (2006).

[4] Daroczy, Z. and Maksa, Gy., Non-negative information functions in Analytic Function Methods in Probability and Statistics. *Colloq. Math. Soc. J. Bolyai 21, Gyires, B., Ed; north Holand: Amesterdam.* (1979), 65-76.

[5] Diderrich, G., The role of boundedness in characterizing Shannon entropy. *Information and Control.* (1975), 29, 140-161.

[6] Fadeev, D.K., On the concept of entropy of a finite probability scheme (in Russian). *Uspehi Mat. Nauk.* (1956), 11, 227-231.

[7] Hartley, R. V. L., Transmission of information. *Bell System Technical Journal.* (1928), 7, 535-563.

[8] Harvda, J. and Charvat, F., Quantification method of classification processes: Concept of structural $\alpha-$entropy. *Kybernetika.* (1967), 3, 30-35.

[9] Koski, T. and Persson, L.E., Some properties of generalized exponential entropies with applications to data compression. *Information Theory.* 50 (1992), 6, 1220-1228.

[10] Lee, P. M., On the axioms of information theory. *Ann. Math. Statist.* (1964) 35, 415-418.

[11] Renyi, A., On measures of entropy and information. Proc. *Berekeley Symposium, Statist. Probability.* (1961), 1, 547-561.

[12] Shannon, C. E., A mathematical theory of communication. *Bell System Technical Journal.*, (1948), 27, 379-423.

[13] Tsallis, C., Possible generalizations of Boltzmann-Gibbs statistics. *Journal of Statistical Physics.* (1988), 52, 479-487.

[14] Tverberg, H., A new derivation of the information function. *Math. Scand.* (1958), 6, 297-298.

Affiliations

G. R. Mohtashami Borzadaran, M. Sanei Tabass, M. Amini
Address: Department of Statistics Faculty of Sciences Ferdowsi University of Mashhad, Mashhad-Iran
e-mail: gmb1334@yahoo.com, grmohtashami@um.ac.ir