RELATION BETWEEN SHANNON ENTROPY, RENYI ENTROPY AND INFORMATION

Rohit Kumar Verma
rohitkverma73@rediffmail.com
Associate Professor, Department of Mathematics
Bharti Vishwavidyalaya, Durg, C.G. India

Jharana Chandrakar
jhar4740@gmail.com
Research Scholar Ph.D. Department of Mathematics
Bharti Vishwavidyalaya, Durg, C.G. India

ABSTRACT

This paper provides evidence that the Renyi Entropy and information are both limiting case of the Shannon Entropy. In addition, experimental evidence supports these findings. As a last thought, we offer some broad conclusions on the usefulness of entropy metrics. In an appendix, a brief history of the idea of physical entropy is given.

INTRODUCTION

Early in the twentieth century, the development of telecommunications inspired a number of scholars to investigate the information content of signals. These early attempts were rationalized into a cogent mathematical theory of communication in Shannon's foundational work, which was based on publication by Nyquist [8,9] and Hartley[6]. This work launched the field of study that is now known as information theory. According to Shannon [12], a measurement of the amount of information S(P) included in the series of events p_1 p_n

must meet three criteria.

- S should be continuous with pi
- If all the $pi$are equally probably, so pi = 1N , then S should be monotonic increasing function of N.
- S should be additive.

He then prove that the only S satisfying these three requirements is

$$S(P) = - M \sum_{i=1}^{N} p_i \ln p_i$$

Where M is a positive constant. Since then, this amount has been referred to as the Shannon entropy. In instance, Shannon entropy is frequently cited as the source of the mutual information measure used in multimodal medical picture registration.

Numerous more measures of information or entropy have been produced as a result of extensions to Shannon's initial study. As an illustration, try loosening the third. Renting was able to expand Shannon entropy to a continuous family of entropy measures that adhere to Shannon's criterion for additivity.

$$S_q(\mathrm{P}) = \frac{1}{1-q} \ln \sum_{i=1}^{N} p_i^q$$

The Renyi [11] Entropy tends to Shannon Entropy as q → 1

In addition Kandell [10] defines the information content of a probability distribution in the discrete as

$$S_q(\mathrm{P}) = \frac{1}{1-q} \ln \sum_{i=1}^{N} p_i^q$$

Which again tends to Shannon Entropy as q → 1

The claims that these expressions regenerate Shannon entropy in the limit have not been supported by any proofs, so we provide them here and test the conclusions experimentally on a sample of uniform probabilities. We end by making a few observations about the general theoretical viability of entropy metrics.

SHANNON ENTROPY AND RENYI ENTROPY

Given a sample of probabilities pi

$$\sum_{i=1}^{N} p_i = 1$$

The Renyi Entropy of the sample is given by

$$S_q(\mathrm{P}) = \frac{1}{1-q} \ln \sum_{i=1}^{N} p_i^q$$

In order to find the limit

$$\lim_{q \to \alpha} \frac{f(q)}{g(q)} = \lim_{q \to \alpha} \frac{f'(q)}{g'(q)}$$

Where in this case α=1. We put

$$\mathrm{f(q)} = \ln \sum_{i=1}^{N} p_i^q \quad \text{and} \quad \mathrm{g(q)} = \frac{1}{1-q}$$

Then $\frac{d\, g(q)}{dq}$ = -1

and applying the chain rule

$$\frac{d\, f(q)}{dq} = \frac{1}{\sum_{i=1}^{N} p_i^q} \sum_{i=1}^{N} \frac{dp_i^q}{dq}$$

SHANNON ENTROPY AND INFORMATION:

The information of sample of probabilities pi, where

$$\sum_{i=1}^{N} p_i = 1$$

is given by

$$K_q\,(\mathrm{P}) = -\sum_{i=1}^{N} \frac{p_i^q}{q-1} + \frac{1}{q-1} = \sum_{i=1}^{N} \frac{p_i - p_i^q}{q-1}$$

Put q-1 = α, so that as q→ 1 *and* $\alpha \rightarrow 0$, and $p_i = 1 - x_i$. Then

$$K_\alpha(\mathrm{X}) = \sum_{i=1}^{N} \frac{(1-x_i) - (1-x_i)^{\alpha+1}}{\alpha}$$

Taking out one power of p_i immediately gives

$$K_\alpha(\mathrm{X}) = \sum_{i=1}^{N} \frac{(1-x_i)[1-(1-x_i)^{\alpha}]}{\alpha}$$

The binomial expansion

$$(1+x)^n = 1 + \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} + \cdots$$

Can be applied to the first term of this equation to give

$$\frac{(1-x_i)^{\alpha} - 1}{\alpha} = -x_i + (\alpha - 1)\frac{x^2}{2!} - (\alpha-1)(\alpha-2)\frac{x^3}{3!} \ldots \ldots ..$$

In the limit of $\alpha \rightarrow 0$ this becomes

$$= -x_i - \frac{x_i^2}{2} - \frac{x_i^3}{3} \ldots \ldots ..$$

Which is the well known series expansion for the natural logarithm

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} \ldots \ldots .$$

Therefore,

$$\lim_{\alpha \to 0} \frac{(1-x_1) - (1-x_i)^{\alpha+1}}{\alpha} = -(1-x_i)\ln(1-x_i)$$

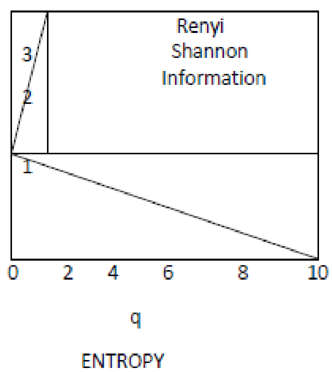
And

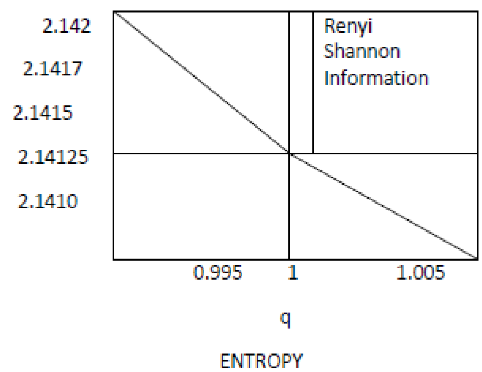
$$K_1\,(\mathrm{P}) = -\sum_{i=1}^{N} p_i \ln p_i$$

Which is the Shannon entropy.

EXPERIMENTAL TESTING:

Plotting Shannon entropy, Renyi entropy and information against q for a sample with uniform probability corroborated the aforementioned findings. A uniform distribution was used to create ten random samples, which were then normalized so that their sum was one. The information and Shannon and Renyi entropies were then plotted against q. The three measures converge as expectedas q approaches towards 1. Around this momentthere has been good behavior.





For a sample of uniform probability with N=10, various entropy metrics . The Shannon Entropy fpr a tends to 1 as the Renyi entropy and information converge . The intersection point in the left hand image is enlarged in the right hand image.

CONCLUSION

This paper has shown that both the Renyi entropy Sq(P) and information Kq(P) tend to the Shannon Entropy in the limit of the q goes to 1. The Renyi entropy is also a monotonic information function. However, when applied to a continuous distribution, as Kendall [10] notes, these metrics are scale dependent, rendering their absolute values useless. As a result, they are typically only appropriate for use in comparative or differential process. Renyi entropy and information can be utilised interchangeably in all practical applications, according to the monotonic connection.

Despite the fact that these entropy measures belong to a family of self consistent functions, their scale dependence restricts their use because they cannot then be regarded as well found statistics. For instance, the mutual information measures that are frequently employed in information theoretic multi modal medical picture coregistration [15] can be derived from concept of Shannon entropy. However, recent research [13,4,1,2,3] has revealed that mutual information is actually a biassed version of maximum likelihood, and that Shannon entropy is the same thing as the likelihood function when used to determine the information content of signals made up of a discrete alphabet of independent symbols. The statistical validity of such procedures would be in doubt even though the Renyi entropy may be utilised to construct a continuous family of mutual information measures that could be applied for example, to coregistration.

REFERENCES:

1. P A Bromiley, M Pokric, and N A Thacker. Computing covariances for mutual information coregistration. In Proceedings MIUA 2004, 2004.
2. P A Bromiley, M Pokric, and N A Thacker. Empirical evaluation of covariance matrices for mutual information coregistration. In Proceedings MICCAI 2004, 2004.
3. P A Bromiley, M. Pokric, and N A Thacker. Tina memo 2004-001: Empirical evaluation of covariance matrices for mutual information coregistration. Technical report, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 2004.
4. P A Bromiley and N A Thacker. Tina memo 2003-002: Computing covariances for mutual information coregistration 2. Technical report, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 2003.
5. A Cho. A fresh take on disorder, or disorderly science? Science, 297:1268–1269, 2002.
6. R V L Hartley. Transmission of information. Bell Systems Technical Journal, page 535, July 1928.
7. J Havrda and F Charvat. Quantification method of classification processes: concept of structural α-entropy.
8. Kybernetika, 3:30–35, 1967. [8] H Nyquist. Certain factors affecting telegraph speed. Bell Systems Technical Journal, page 324, April 1924.
9. H Nyquist. Certain topics in telegraph transmission theory. A.I.E.E. Trans., page 617, April 1928.
10. K Ord and S Arnold. Kendall's Advanced Theory of Statistics: Distribution Theory. Arnold, 1998.
11. A Renyi. On measures of entropy and information. In Proc. Fourth Berkeley Symp. Math. Stat. Prob., 1960, volume 1, page 547, Berkeley, 1961. University of California Press.

12. C E Shannon. A mathematical theory of communication. Bell Systems Technical Journal, 27:379–423 and 623–656, Jul and Oct 1948.
13. N A Thacker and P A Bromiley. Tina memo 2001-013: Computing covariances for mutual information coreg- istration. Technical report, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 2001.
14. C Tsallis. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics, 52:479–487, 1988. P Viola andWMWells. Alignment by maximisation of mutual information. International Journal of Computer Vision, 24(2):137–154, 1997.