

Shannon Entropy, Renyi Entropy, and Information

P.A. Bromiley, N.A. Thacker and E. Bouhova-Thacker

Last updated
17 / 9 / 2004

This document forms part of the **Statistics and Segmentation Series (2008-001)**
available from www.tina-vision.net.

- 2007-008 Tutorial: Defining Probability for Science.
- 2001-007 Performance Characterisation in Computer Vision:
The Role of Statistics in Testing and Design.
- 2002-007 The Effects of an Arcsin Square Root Transform on a Binomial Distributed Quantity.
- 2001-010 The Effects of a Square Root Transform on a Poisson Distributed Quantity.
- 2004-004 Shannon Entropy, Renyi Entropy, and Information.
- 2002-002 Validating MRI Field Homogeneity Correction Using Image Information Measures.
- 2004-001 Empirical Validation of Covariance Estimates for Mutual Information Coregistration.
- 2004-005 The Equal Variance Domain: Issues Surrounding the Use of Probability Densities in
Algorithm Design.
- 2009-008 Avoiding Zero and Infinity in Sample Based Algorithms.
- 2001-008 Derivation of the Renormalisation Formula for the Product of Uniform Probability
Distributions and Extension to Non-Integer Dimensionality.
- 2001-005 Model Selection and Convergence of the EM Algorithm.
- 2003-007 Noise Filtering and Testing for MR Using a Multi-Dimensional Partial Volume Model.
- 2002-004 A Novel Method for Non-Parametric Image Subtraction:
Identification of Enhancing Lesions in Multiple Sclerosis from MR Images.
- 2001-014 Bayesian and Non-Bayesian Probabilistic Models for Image Analysis.
- 1997-001 The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.
- 1999-001 The Bhattacharyya Measure requires no Bias Correction.
- 1999-004 B-Fitting: An Estimation Technique With Automatic Parameter Selection.
- 2005-008 Tutorial: Beyond Likelihood.



Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
Stopford Building, The University of Manchester,
Oxford Road, Manchester M13 9PT, U.K.

Shannon Entropy, Renyi Entropy, and Information

P.A. Bromiley, N.A. Thacker and E. Bouhova-Thacker.
Imaging Science and Biomedical Engineering,
School of Cancer and Imaging Sciences,
Stopford Building, The University of Manchester,
Oxford Road, Manchester M13 9PT, U.K.
email: paul.bromiley@man.ac.uk

Abstract

This memo contains proofs that the Shannon entropy is the limiting case of both the Renyi entropy and the information. These results are also confirmed experimentally. We conclude with some general observations on the utility of entropy measures.

1 Introduction

The growth of telecommunications in the early twentieth century led several researchers to study the information content of signals. The seminal work of Shannon [10], based on papers by Nyquist [6, 7] and Hartley [5], rationalised these early efforts into a coherent mathematical theory of communication and initiated the area of research now known as information theory. Shannon states that a measure of the amount of information $H(p)$ contained in a series of events $p_1 \dots p_N$ should satisfy three requirements:

- H should be continuous in the p_i ;
- if all the p_i are equally probably, so $p_i = 1/N$, then H should be a monotonic increasing function of N ;
- H should be additive.

He then proved that the only H satisfying these three requirements is

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

where K is a positive constant. This quantity has since become known as the Shannon entropy. It has been used in a variety of applications: in particular, Shannon entropy is often stated to be the origin of the mutual information measure used in multi-modality medical image coregistration.

Extensions of Shannon's original work have resulted in many alternative measures of information or entropy. For instance, by relaxing the third of Shannon's requirements, that of additivity, Renyi [9] was able to extend Shannon entropy to a continuous family of entropy measures that obey

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

The Renyi entropy tends to Shannon entropy as $q \rightarrow 1$.

In addition, Kendall [8] defines the information content of a probability distribution in the discrete case as

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

which again tends to the Shannon entropy as $q \rightarrow 1$.

We have not been able to find proofs for the assertions that these expressions regenerate Shannon entropy in the limit, and we therefore present such proofs here, and confirm the results experimentally on a sample of uniform probabilities. We conclude with some observations on the theoretical validity of entropy measures in general.

2 Shannon Entropy and Renyi Entropy

Given a sample of probabilities p_i

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

the Renyi entropy of the sample is given by

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

At $q = 1$ the value of this quantity is potentially undefined as it generates the form $0/0$. In order to find the limit of the Renyi entropy, we apply l'Hopital's Theorem

$$\lim_{q \rightarrow a} \frac{f(q)}{g(q)} = \lim_{q \rightarrow a} \frac{f'(q)}{g'(q)}$$

where in this case $a = 1$. We put

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

Then

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

and, applying the chain rule

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

The form a^x can be differentiated w.r.t. x by putting

$$\frac{d}{dx} a^x = \frac{d}{dx} e^{x \ln a} = e^{x \ln a} \frac{d}{dx} x \ln a = a^x \ln a$$

Therefore

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

Letting $q \rightarrow 1$, we have

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

collecting terms we have

$$\lim_{q \rightarrow a} \frac{1}{1-q} \ln \sum_{i=1}^N p_i^q = - \sum_{i=1}^N p_i \ln p_i$$

which is the Shannon entropy.

3 Shannon Entropy and Information

The information of a sample of probabilities p_i where

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

is given by

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

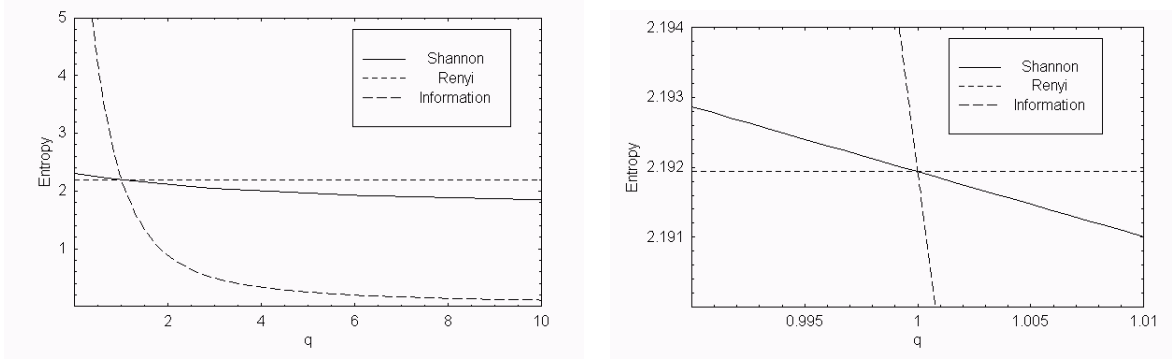


Figure 1: Various entropy measures for a sample of uniform probabilities with $N = 10$. The Renyi entropy and information converge to the Shannon entropy for $q \rightarrow 1$.

Put $q - 1 = a$, so that as $q \rightarrow 1$ $a \rightarrow 0$, and $p_i = 1 - x_i$. Then

$$I_a(X) = \sum_{i=1}^N \frac{(1 - x_i) - (1 - x_i)^{a+1}}{a}$$

Taking out one power of p_i immediately gives

$$I_a(X) = \sum_{i=1}^N \frac{(1 - x_i)(1 - (1 - x_i)^a)}{a}$$

The binomial expansion

$$(1 + x)^n = 1 + nx + n(n-1)\frac{x^2}{2!} + n(n-1)(n-2)\frac{x^3}{3!} \dots$$

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

$$\frac{(1 - x_i)^a - 1}{a} = -x_i + (a-1)\frac{x_i^2}{2!} - (a-1)(a-2)\frac{x_i^3}{3!} \dots$$

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

$$= -x_i - \frac{x_i^2}{2} - \frac{x_i^3}{3} \dots$$

Which is the well known series expansion for the natural logarithm

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

Therefore,

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

and

$$I_1(X) = - \sum_{i=1}^N p_i \ln p_i$$

which is the Shannon entropy.

4 Experimental Testing

The above results were confirmed by plotting the Shannon entropy, Renyi entropy, and information against q for a sample of uniform probabilities. Ten random samples from a uniform distribution were generated and normalised such that they summed to unity. Then the Shannon and Renyi entropies and information were plotted against q . The results are shown in Fig. 1. As expected, the three measures converge as $q \rightarrow 1$. The behaviour around this point is well behaved.

5 Conclusions

This memo has demonstrated that, in the limit of $q \rightarrow 1$, both the Renyi entropy $H_q(p)$ and the information $I_q(p)$ tend to the Shannon entropy. Also, the Renyi entropy is a monotonic function of the information. However, as Kendall states [8] these measures are scale-dependent when applied to continuous distributions, and so their absolute values are meaningless. Therefore, they can generally only be used in comparative or differential processes. The monotonic relationship therefore implies that Renyi entropy and information can be used interchangeably in any practical applications.

Although these entropy measure form a self-consistent family of functions, their scale-dependence limits their utility as they cannot then be considered as well-formed statistics. For instance, concepts of Shannon entropy can be used to derive the mutual information measure commonly used in information-theoretic multi-modality medical image coregistration [12]. However, recent work [11, 4, 1, 2, 3] has shown that mutual information is in fact a biased maximum likelihood technique, and in the original application of Shannon entropy, calculating the information content of signals composed from a discrete alphabet of independent symbols, Shannon entropy is identical to the likelihood function. Therefore, although the Renyi entropy could be used to derive a continuous family of mutual information measures that could be applied, for instance, to coregistration, the statistical validity of such techniques would be questionable.

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] R V L Hartley. Transmission of information. *Bell Systems Technical Journal*, page 535, July 1928.
- [6] H Nyquist. Certain factors affecting telegraph speed. *Bell Systems Technical Journal*, page 324, April 1924.
- [7] H Nyquist. Certain topics in telegraph transmission theory. *A.I.E.E. Trans.*, page 617, April 1928.
- [8] K Ord and S Arnold. *Kendall's Advanced Theory of Statistics: Distribution Theory*. Arnold, 1998.
- [9] 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.
- [10] C E Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423 and 623–656, Jul and Oct 1948.
- [11] N A Thacker and P A Bromiley. Tina memo 2001-013: Computing covariances for mutual information coregistration. Technical report, Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester, 2001.
- [12] P Viola and W M Wells. Alignment by maximisation of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.