Entropy*

Anders Gjendemsjø
Behnaam Aazhang

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 1.0†

Abstract

Entropy

1

The self information gives the information in a single outcome. In most cases, e.g in data compression, it is much more interesting to know the **average information content** of a source. This average is given by the **expected** value of the self information with respect to the source's probability distribution. This average of self information is called the source entropy.

1.1 Definition of entropy

Definition 1: Entropy
1. The entropy (average self information) of a discrete random variable X is a function of its probability mass function and is defined as

$$H(X) = -\sum_{i=1}^{N} p_X(x_i) \log p_X(x_i) \tag{1}$$

where N is the number of possible values of X and $P_X(x_i) = Pr[X = x_i]$. If log is base 2 then the unit of entropy is bits per (source)symbol. Entropy is a measure of uncertainty in a random variable and a measure of information it can reveal.
2. If symbol has zero probability, which means it never occurs, it should not affect the entropy. Letting $0 \times \log 0 = 0$, we have dealt with that.

In texts you will find that the argument to the entropy function may vary. The two most common are $H(X)$ and $H(p)$. We calculate the entropy of a source X, but the entropy is, strictly speaking, a function of the source's probabilty function p. So both notations are justified.

*Version 1.4: May 18, 2004 1:39 pm -0500
†http://creativecommons.org/licenses/by/1.0

1.2 Calculating the binary logarithm

Most calculators does not allow you to directly calculate the logarithm with base 2, so we have to use a logarithm base that most calculators support. Fortunately it is easy to convert between different bases.

Assume you want to calculate $\log_2 x$, where $x > 0$. Then $\log_2 x = y$ implies that $2^y = x$. Taking the natural logarithm on both sides we obtain

NOTE: $\log_2 x = \frac{\ln(x)}{\ln(2)}$

1.3 Examples

Example 1
When throwing a dice, one may ask for the average information conveyed in a single throw. Using the formula for entropy we get $H(X) = -\sum_{i=1}^{6} p_X(x_i) \log p_X(x_i) = \log 6 \text{bits/symbol}$

Example 2
If a soure produces binary information $\{0, 1\}$ with probabilities p and $1 - p$. The entropy of the source is

$$H(X) = (-(p\log_2 p)) - (1-p)\log_2(1-p) \tag{2}$$

If $p = 0$ then $H(X) = 0$, if $p = 1$ then $H(X) = 0$, if $p = 1/2$ then $H(X) = 1$. The source has its largest entropy if $p = 1/2$ and the source provides no new information if $p = 0$ or $p = 1$.

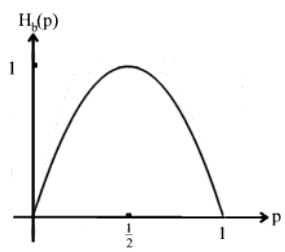

Figure 1

Example 3
An analog source is modeled as a continuous-time random process with power spectral density bandlimited to the band between 0 and 4000 Hz. The signal is sampled at the Nyquist rate. The sequence of random variables, as a result of sampling, are assumed to be independent. The samples are quantized to 5 levels $\{-2, -1, 0, 1, 2\}$. The probability of the samples taking the quantized values are $\left\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\right\}$, respectively. The entropy of the random variables are

$$\begin{aligned} H(X) &= -\sum_{i=1}^{5} p_X(x_i) \log p_X(x_i) \\ &= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{1}{4} + \frac{1}{4} \\ &= \frac{15}{8} \text{bits/sample} \end{aligned} \tag{3}$$

There are 8000 samples per second. Therefore, the source produces $8000 \times \frac{15}{8} = 15000$ bits/sec of information.

Entropy is closely tied to source coding. The extent to which a source can be compressed is related to its entropy. There are many interpretations possible for the entropy of a random variable, including

- (Average)Self information in a random variable
- Minimum number of bits per **source symbol** required to describe the random variable without loss
- Description complexity
- Measure of uncertainty in a random variable

2 References

- Øien, G.E. and Lundheim,L. (2003) **Information Theory, Coding and Compression**, Trondheim: Tapir Akademisk forlag.