S6: Connecting Hartley Measure, Shannon Entropy and AEP

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Scribe: 6

Student: Pranjal Umesh Kalekar

CS 7840: Foundations and Applications of Information Theory (fa24)

https://northeastern-datalab.github.io/cs7840/fa24/

Lecturers: Wolfgang Gatterbauer, Javeed Aslam

Dec 11, 2024

Quantifying Uncertainty: The Hartley Measure and Beyond

The Hartley measure quantifies the amount of uncertainty associated with a finite set of possible alternatives. It is based on the idea that uncertainty **increases** with the number of possibilities.

<u>Derivation:</u>

Consider a finite set X of possible alternatives, where only one is the true alternative.

Evidence suggests that only alternatives within a subset E of X are possible.

Hartley proposed that uncertainty associated with E, denoted by H(E), should be a function of the number of elements in E (its cardinality, |E|).

Logarithmic Function: The most intuitive way to measure this uncertainty is using a logarithmic function because:

Additivity: The uncertainty of two independent events should be the sum of their individual uncertainties. Logarithms exhibit this property: log(a*b) = log(a) + log(b).

Monotonicity: Uncertainty increases with the number of possibilities. Logarithmic functions are monotonically increasing.

From Intuition to the Hartley Measure Formula

This leads to the general form of the Hartley measure:

$$H(E) = c * \log_b(|E|),$$

where b and c are parameters ($b > 1$, $c > 0$) defining the measurement unit.

Choosing $b = 2$ and $c = 1$ results in measuring uncertainty in bits.

One bit represents the uncertainty of choosing between **two** equally likely alternatives.

The Hartley measure in bits is:

$$H(E) = \log_2(|E|)$$

Embracing Probabilities with Shannon Entropy

The Shannon entropy function generalizes the concept of uncertainty to probability distributions. It quantifies the average uncertainty associated with a random variable.

Consider a set of events with probabilities of occurrence $p_1, p_2, ..., p_n$. The goal is to find a measure, $H(p_1, p_2, ..., p_n)$, for the "choice" involved or the uncertainty of the outcome.

Shannon proposed several intuitive properties for this measure:

- Continuity: H should be continuous in the p_is.
- Monotonicity: If all p_is are equal (1/n), H should increase with n.
- Decomposability: If a choice is broken down into successive choices, the original H should be the weighted sum of the individual H values.

These properties uniquely determine the entropy function (up to a constant factor):

$$H = -K * \sum(p_i * \log(p_i)),$$

where K is a positive constant that sets the unit of measurement.

Choosing K = 1 and using the base-2 logarithm results in measuring entropy in bits.

The Shannon entropy in bits is:

$$H(X) = - \sum(p_i * \log_2(p_i))$$

Connecting Hartley's measure and Shannon's entropy to Compression with the AEP

Hartley Measure as a Foundation: The Hartley measure, $H(E) = \log_2(|E|)$, provides a basic way to quantify the uncertainty associated with choosing from a finite set of <u>equally</u> likely alternatives. It forms a conceptual steppingstone to the more general Shannon entropy.

Shannon Entropy Generalizes Uncertainty: The Shannon entropy extends the concept of uncertainty to probability distributions, considering the likelihood of each event. Its formula, $H(X) = - \sum(p_i * \log_2(p_i))$, captures the average uncertainty of a random variable.

AEP Connects Entropy to Typical Sequences: AEP states that for a large enough sequence length n, the probability mass concentrates in a subset called the "typical set," Aϵ(n).

Typical Set Size and Entropy: The size of this typical set is roughly $2^{(n*H(U))}$, where H(U) is the Shannon entropy of the source. This connection is significant because it shows that the <u>entropy dictates the effective number of sequences that are likely to occur.</u>

AEP and Compression Efficiency

- The AEP demonstrates that efficient compression schemes can be designed by focusing on encoding the typical sequences, which carry most of the probability mass. Trying to represent sequences outside the typical set becomes increasingly inefficient as n grows.
- **Near-Lossless Compression**: This leads to the idea of "near-lossless" compression, where non-typical sequences, with vanishingly small probabilities, are disregarded to achieve a high compression rate.
- **Lower Bound on Compression**: AEP also establishes a lower bound on the number of bits per symbol required for near-lossless compression, which is **H(U)**. This means that any fixed-length compression scheme using fewer than **H(U)** bits per symbol will fail to represent the input sequence with high probability as n increases.

References:

- https://stanforddatacompressionclass.github.io/notes/lossless_iid/aep.html, Asymptotic Equipartition Property, EE274: Data Compression, course notes
- https://www.tandfonline.com/doi/epdf/10.1080/03081079.2010.526757?needAccess=true, A note on the Hartley-like measure of uncertainty, George J. Klir, Feb 2011
- https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf, A Mathematical Theory of Communication, C. E. SHANNON