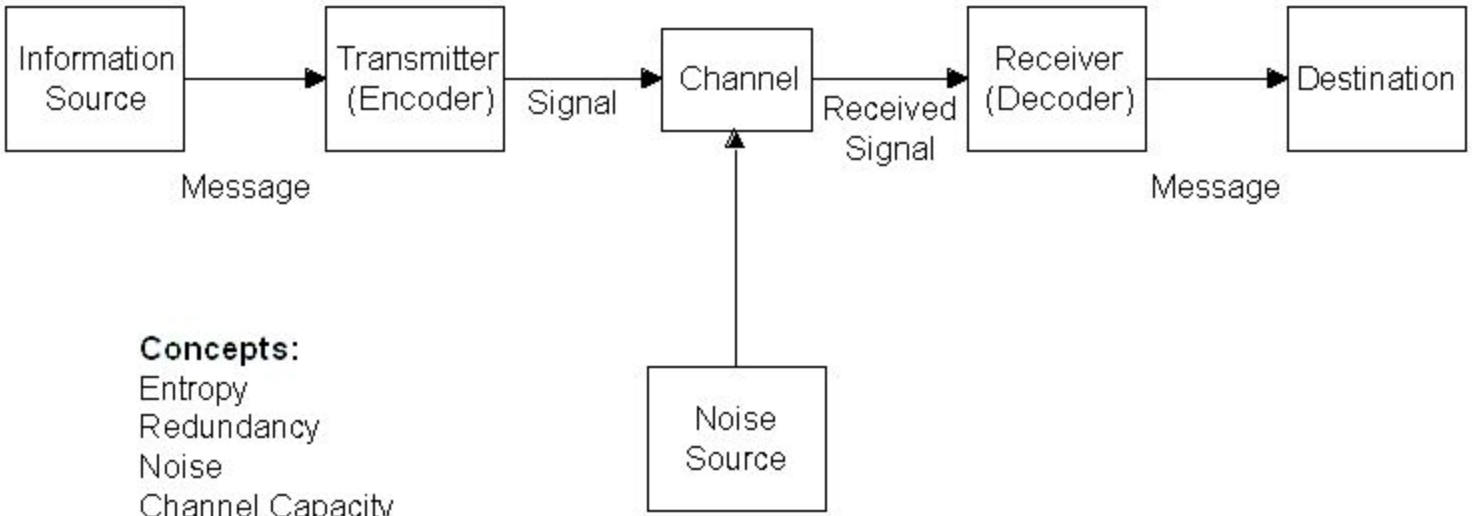
Information Theory + Compression

Benjamin Garcia

Overview of Information Theory

- Relatively new discipline
  - “A Mathematical Theory of Communication” (Shannon 1948)
- Broadly speaking, deals with:
  - Transmission of data through mechanisms of encoding (compression), decoding, mathematical/probabilistic analysis
- DOES NOT DEAL WITH:
  - SEMANTICS
- We’ll go over the origins of information theory
  - Lots of Shannon, 1948
  - A little Von Neumann

The Shannon-Weaver Mathematical Model, 1949



Concepts:
Entropy
Redundancy
Noise
Channel Capacity

Shannon-“A Mathematical Theory of Comm..”

- A monster
- Shannon idea of information:
  - Semantics are irrelevant regarding communication -- “The significant aspect is that the actual message is one selected from a set of possible messages” (MTC 1)
    - Idea: increased information comes from a message’s *not* being another potential message
- Information can take on many forms
  - Sequence of characters (discrete)
  - f(t)
  - $f(t,x_1,...,x_n)$
  - f(t), g(t), …
  - other combinations of above

Shannon intro

- Shannon def. of information immediately gives way to measure of information:
  - **Naive metric of information = $\log_2 M$ bits** where M is the size of the set of possible messages/symbols (in discrete case)
  - This set can be: letters, words, roman numerals, etc.
  - Example for arabic numerals: $\log_2 10 = 3.32$ bits, so each base 10 number can be expressed in 3.32 bits
- This will be refined soon
- We'll be focusing on the discrete case -> don't worry about functions and integrals and stuff

Channel Capacity

- **Maximum rate of bits/second transmittable by channel**
- **C=** (lim T->infinity)(log(N(T))/T)
  - where N(T) is the number of potential signals of duration T
  - From now on, baseless log always means log base 2

Probabilistic Information Sources

Intuition- in natural English, the letter “e” occurs more frequently than the letter “q”

In the same way, the word “and” is more frequent than “vociferous”

You can assign each element of the set of possibly transmitted symbols a probability

This generates a probability mass function which characterizes the nature of the information source -- so the information source is a random variable X

After this Shannon describes a lot of conditional probability and Markov modeling of information sources in depth- this we will skip

Background to Entropy

Now our source is a discrete RV X.

We want a metric such that we can measure the rate of information produced from this random process.

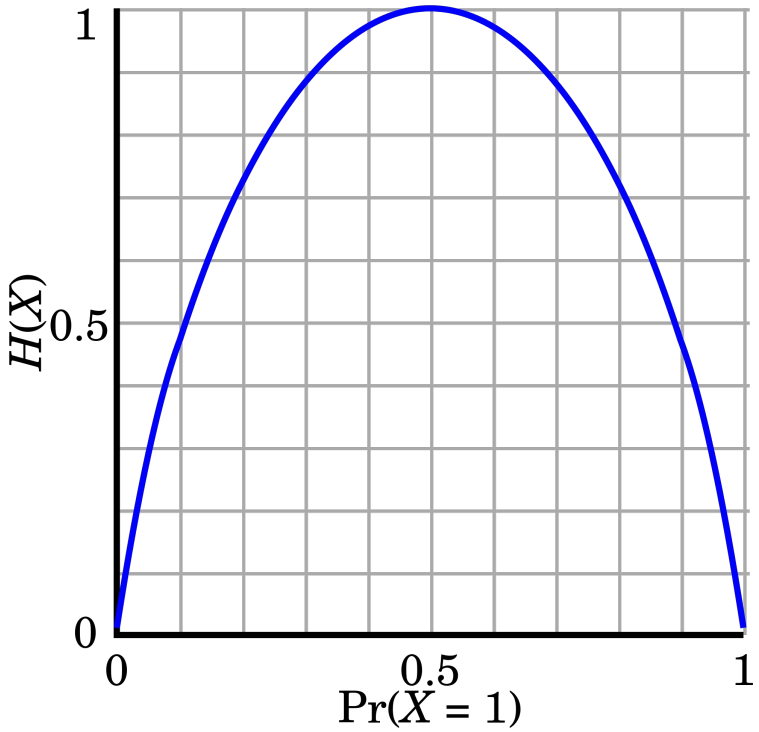
Say $p(X = x_n) = a_n$, and our new metric is H(X)

H(X) is the negative sum over all possible values of X of $p_i \log(p_i)$

This is called “entropy”, due to a similar mathematical/statistical formulation as thermodynamic entropy

Unit of H(X): bit/symbol

H(X) of a binary generator



Intuition for Entropy

It's exactly like thermodynamic entropy: a measure of certainty/order

Information gleaned from a symbol is related to its probability of being any symbol

Shannon's 3 intuitive conditions when defining entropy:

H(X) where X takes on a pmf $\{p_1,p_2,...,p_n\}$

1. H is continuous in the p_i
2. If all p_i are evenly distributed on X, H increases with n
3. If a choice denoted p_i is separated into 2 sub-choices, the total H is representable as a weighted sum of the individual values of H

Joint entropy

Let p(i,j) be the joint PMF for 2 R.V.s X and Y

Then

$$H(x, y) = -\sum_{i,j} p(i, j) \log p(i, j)$$

and

$$H(x) = -\sum_{i,j} p(i, j) \log \sum_{j} p(i, j)$$

$$H(y) = -\sum_{i,j} p(i, j) \log \sum_{i} p(i, j).$$

Other relations to probability

Conditional Probability -> Conditional Entropy

$$H_x(y) = -\sum_{i,j} p(i, j) \log p_i(j).$$

Old notation: H(Y given X) -> $H_X(Y)$

Complete relational inequality:

$$H(x) + H(y) \geq H(x, y) = H(x) + H_x(y)$$

Entropy of a Source

If each letter in our string is generated by a random variable on a set of letters, successive letters are not necessarily independent AND not necessarily generated by the same R.V.

If they aren't independent, you have entropy H_i for each state

Then your expected entropy is

negative sum over i of $P_i H_i$

where P is the probability of the specific random variable generating in the sequence, and H is the entropy of the random variable

Calculable indirectly using laws of large numbers, limiting processes

Redundancy

Relative entropy = H/(max H using same # of symbols)

Redundancy = 1-R.E.

Redundancy takes values on [0,1]

Measures the percent of symbols which could be taken out of the signal without losing information

Hard bound on lossless compression under this model of information

Fundamental Theorem for Noiseless Channels

More on the hard bound:

Let the source have an entropy of H(bits/symbol) and the channel have a capacity of C (bits/second)

Then **it is possible to encode (compress) the source so as to transmit at C/H symbols/second**

Example

A source chooses from {A, B, C, D} independently w/ p(A)= ½, p(B) = ¼, p(C) and p(D) = ⅛

H= 7/4 bits/symbol

Current information density= log(4) = 2 bits

Code: A=0, B=10, C= 110, D= 111

Then average bits in encoding N symbols will be 7/4 bits/symbol

(½*1+¼*2+⅛*3+⅛*3)=7/4 bits, as opposed to 2 bits

Von Neumann Entropy (quantum)

If a quantum system is described by a matrix ρ on a m-dimensional complex Hilbert space C^n, its Von Neumann entropy (S) is

$$S = -\text{trace}(\rho \ln \rho)$$

then if ρ is decomposed to eigenvectors λ_i, S is rewritable as

$$S = \text{negative sum over all } i \ (\lambda_i \ln \lambda_i)$$

which is analogous to Shannon entropy

The value of S changes with change of basis on C^n!

Quantum Compression Scheme

Classical compression scheme: invertible function s.t.

$f: \{0,1\}^n \rightarrow \{0,1\}^n$

Quantum: unitary change-of-basis transformation

$QC: C^{2n} \rightarrow C^{2n}$

In Conclusion

Entropy is a measure of uncertainty: the more uncertain you are that a symbol will be generated, the more information that symbol carries

Under Shannon theory, channel capacity and information source entropy interact to let you know how fast you can transmit information w/ compression

Shannon theory is applicable to quantum computing, but shifts everything from sets of symbols to Hilbert spaces

There is far more information in the field: to name 3 essential things, noisy signals, continuous signals, and Kolmogorov complexity