Entropy
CSCI 384: Computational Linguistics

CSCI 384

October 11, 2017

Example

The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate H(X), H(Y), H(X|Y), and H(Y|X).

The meaning of entropy

The word *entropy* had of course been used before Shannon. In 1864 Rudolf Clausius introduced the term... to represent a "transformation" that always accompanies a conversion between thermal and mechanical energy. ...

[One of the authors] asked Shannon what he had thought about when he had finally confirmed his famous measure. Shannon replied: "My greatest concern was what to call it. I thought of calling it 'information,' but that word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it *entropy*, for two reasons. In first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy is, so in a debate you will always have the advantage.' "

Tribus and McIrvine, "Energy and Information", *Scientific American* # 224, Sept 1971, pg 178–184

Perplexity and cross entropy

We suspect that speech recognition people prefer to report on the larger non-logarithmic numbers given by perplexity mainly because it is much easier to impress funding bodies by saying that "we've managed to reduce preplexity from 950 to only 540" than by saying that "we've reduced cross entropy from 9.9 to 9.1 bits." However, perplexity does also have an intuitive reading: a perplexity of k means that you are as surprised on average as you would have been if you had had to guess between k equiprobable choices at each step.

Manning and Schütze, *Foundations of Statistical Natural Language Processing*, pg 78.