Understanding Entropy

Shuai

The original intuitive is the smaller the probability one event may happen, the more its uncertainty is. This make sense. It converts the unmeasurable into the measurable.

Now the problem becomes what is the relation between the probability of events and the overall uncertainty of the state?

One state may contain several events, which each of them have their own probability to happen. Using the terminology from Thermodynamics, one microstates could consist of a number of states. Returning back to the probability, it means one state contains a number of events. To take into account the weight of different events, the probabilities are used.

See the p_i in: $entropy = \Sigma(p_i * log(p_i))$

Now, the weight part has done, how can we measure uncertainty?

In the book *Warmth Disperses and Time Passes*, page 106, writen by Hans Christian Von Baeyers:

> Whenever you multiply two integers, the numbers of their respective digits add.
>
> eg: 60 x 600 = 36,000
>
> so two digits plus three digits equals five digits (the rule sometimes misses by one digit, as in 3x3 = 9, but that's a negligible error in view of the vastness of the number of molecules in a gas.)
>
> So Boltzmann made the bold, inspired guess that entropy equals the number of digits of the corresponding probability

I do not read this book. The material comes from this url.

So, this solved my long puzzled question why log is used. We have to think of a way to measure the uncertainty in one event. And such measurement should be mathematically correct. That means, it should satisfy several criteria(and obviously I do not know exactly what are them all):

1. If only one event is contained in the state, the entropy should be 0.
2. If all events happen equally, the entropy should be maximum.
3. ...

Logarithm does a great job. But what is the theory underlying it? Keep reading.

0.1 Another Example

I also tried to understand entropy from encoding perspecitive, by comparing average encoding length(AEL) with entropy.

The AEL is computed as following:

$$AEL = \Sigma(p_i * l_i)$$

Where P_i stands for the probability of one information i(can be string, etc.) being chosen, and l_i stands for the length of the encoding of the information.

In this equation, l_i is semantically similar to the corresponding $\log(p_i)$ part of the Shannon entropy.

Huffman encoding is a perfect example combine encoding with entropy(The detail of Huffman encoding is omitted).

The main intuitive is the longer the encoding of one information, the smaller the probability it may happen. The smaller the AEL of one encoding method is, the better, meaning less chaotic, it is.

Combined with the AEL formula above, the length of the encoding of one information is actually similar to the number of digitals the probability have. And, this connects the dots.