Probability and Statistics (MA6.101)

Monsoon 2021, IIIT Hyderabad
17 September, Friday (Lecture 8)

Taught by Prof. Pawan Kumar

Information Theory

Entropy (in physics) is defined as the log of the number of microstates and microscopic configurations.

Consider a bucket with n balls – r of them red and b blue – in it. We are shown the balls in some order. Now, we draw one ball out and put it back n times. What is the probability p that we get them in the same order again? Clearly

$$p = \left(\frac{r}{n}\right)^r \left(\frac{b}{n}\right)^b.$$

For example, let $n = 4$. If $r = 4$, we get $p = 1$; if $r = 3$, then $p = 0.105$; and if $r = b = 2$, then $p = 0.0625$.
We want a measure of entropy of the bucket that attains a maximum at $r = b$.
Consider $-\frac{1}{n}\log_2 p$ (normalised to make it ≤ 1).

Thus, the general formula for entropy is

$$-\sum_{i=1} np_i \log_2 p_i$$

where n is the number of classes and p_i is the probability of an object from the i^{th} class appearing.

This expresses the number of questions we need on average to find out which letter we have.

Consider a bucket with n letters. If the contents are $AAAAAAAA$, then we need 0 questions; if $AAAABBCD$, then we can ask 3 questions (A? B? C? D?) which means we need
1 question for 4 of the letters,
2 questions for 2 of the letters,
3 questions for 2 of the letters,
which makes the average 1.75.

In case we have $AABBCCDD$, we can ask 2 questions (A or B?: if yes, then A? B?; if no, then C? D?). Then we need 2 questions for all the letters, making the average also 2.