Will you believe if I say log has memory to depict the loss? How does a log of a value give this meaning? I was wondering with examples and want to show you how log can blow our minds.

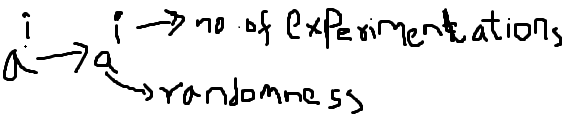
Loss:

It says if our work did not reach to the expectation. So, to reach the expectation we will need something more certain. Being more certain is more important and it is decided to the space or range you are put into.

Ex: Let’s take 6-sided cube and assume we have number 4 in one of the sides. So, when you roll the dice, you could be more certain if you have more 4’s occupying the sides or a smaller number of faces.

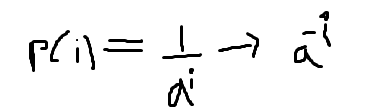
So, the first take away is the more probability you get for the expected, the more certain you will be, the lesser you are subjugated to loss.

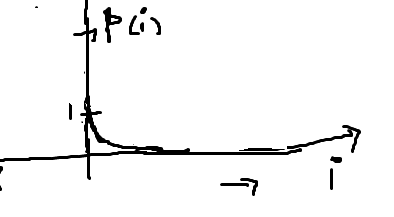
I am very interested to consider the certainty of the expectation, if you are opened to more randomness (meaning a greater number of faces in the dice in the example). Let ‘a’ be number of faces.

The greater number of sides ‘a’ the more unlikely you get the expected number.

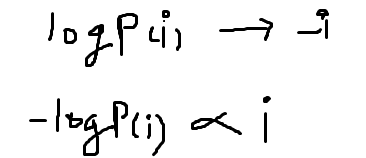
The more number experimentations you do based on the number of faces, your likeliness to the expected depends.

That is the reason, the combined effect is always taken, then you define Probability p(i):

Probability is high if the certainty to the expected is high and low vice versa. The low probability the more you consider your loss.

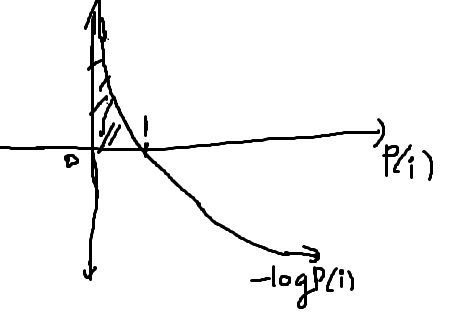
It shows how the probability value keeps dropping down with increasing i. Generally, the graph shows for negative i values and I cannot consider it.

I would like to take further steps to understand the kind of linear combination of the i values with respective probabilities as probabilities are being dropped so easily with initial i values. Then the idea comes to take log



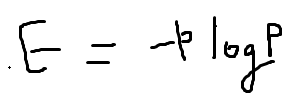
What do we understand by the above steps. There are two important observations you can derive:

1. You are bringing the relation between number of experimentations and the probability. This also defines that the more experimentations you are subjected to you prefer more loss. That is the reason -logP holds loss information
2. Because of the log transformation, you can now be able to catch the loss result with the growing experimentations or randomness/reduced probabilities.

The highlighted area is the transformed part where your probability tends to move to zero, the log property moves higher and higher.

Now the probability above tends to 1, loss goes to 0 because you are able to reach expectation 100 percent. That is the key take away.

That is the reason Entropy might have defined as combined effect of the loss and probability reached. It is expressed as:

where E is entropy of the system.