Justin Marion

Prof Long

CSE 13S

May 9, 2021

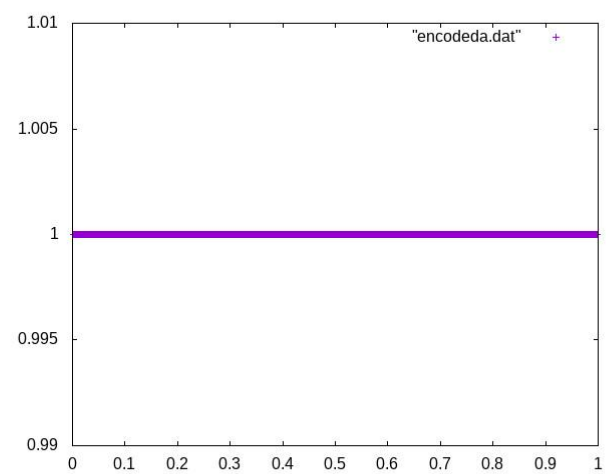
Asgn 5 Writeup

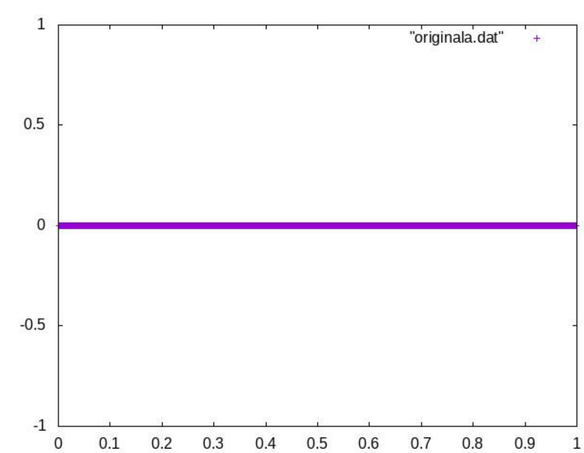
In this lab we created a hamming code encoder. We encode a file and then decode it. The decoder will read the parity bits put into place from the encoder and will be able to detect and correct errors in the bits. Using entropy we can analyze this further.

In this lab we analyzed the change in entropy when a file has more or less error. Here are the graphs I was able to produce from the following example texts given to us in resources:

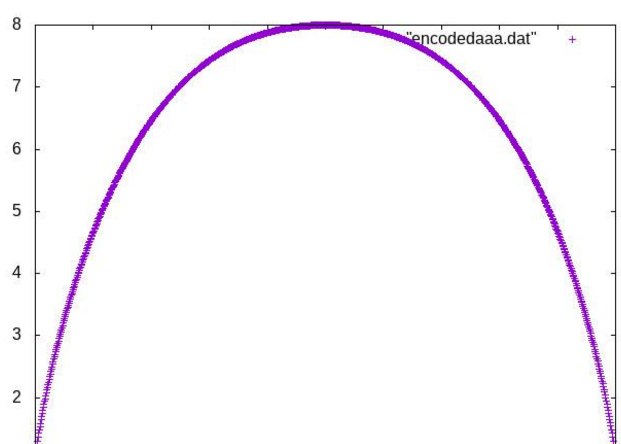
In these graphs we can see the entropy level on the y axis while the error rate on the x axis

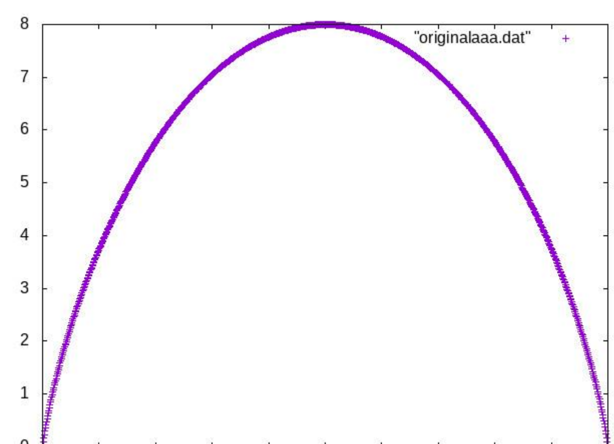
A.txt:



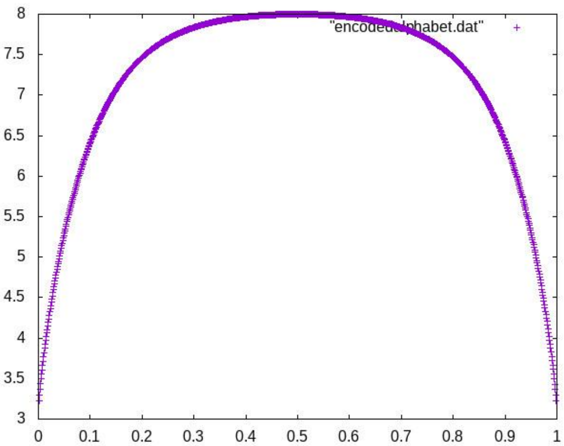


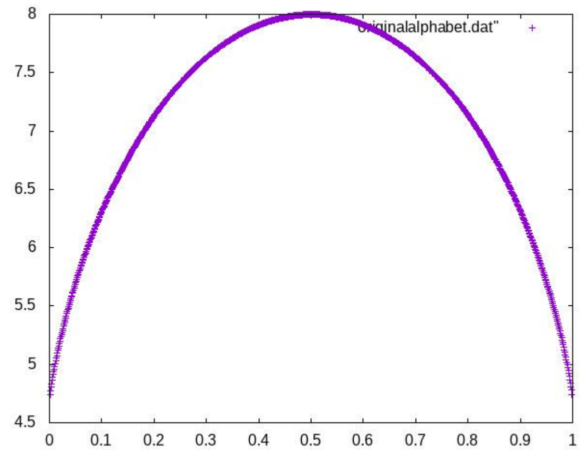
Aaa.txt:



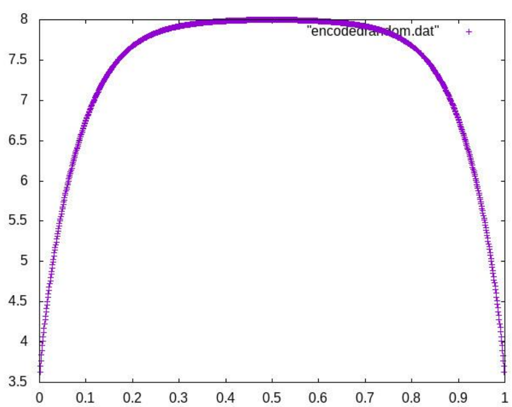


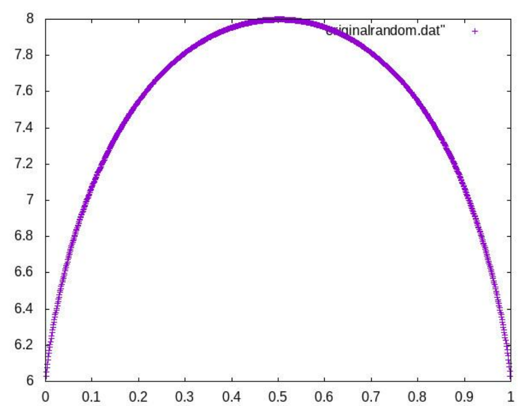
Alphabet.txt:





Random.txt:





Let's first analyze a.txt. This is the simplest to analyze as we can see the entropy does not change at all. Why could this be? This is because it is one simple character, it is very hard to gather data on one simple data point where an error rate can not be well represented. So let's move on to the others.

All the other graphs are extremely similar so let's analyze them together. We can see that they are all bell like curves which is very interesting, let's address this later. A key characteristic that is important to observe is how all the encoded entropy graphs seem to have a higher level of entropy quicker than the original text files. Why is this? This is because as we introduce more “noise” to our data we can expect to be more and more uncertain in our data.Another significant trend that we can see is that the entropy goes back down to introduce an error rate of 100%. This is very interesting, but why is this happening? This is because as we get closer and close to an error rate of 100%, or bits are no longer randomly incorrect. They are simply all flipped. Take this for example:

01011010001010110

If we randomly flip a 4 bits in this array of bits we are going to have a high level of entropy as we become more and more uncertain of the data provided. Hower, we can say that that if we implement an error rate of 100% we can be certain that the output will be:

10100101110101001

The exact opposite of the example above and therefore our entropy goes back down as our error rate approaches 100%.