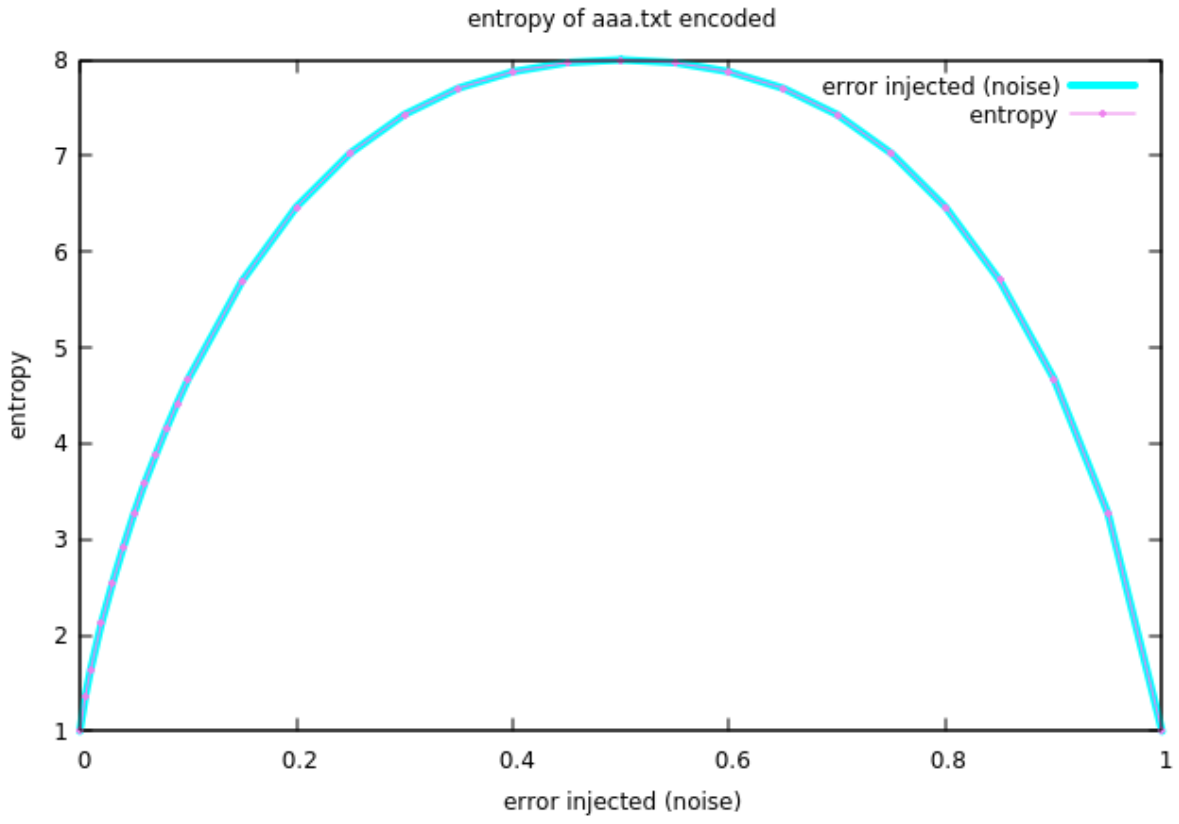
Merilyn Kuo
CSE 13S
Spring 2021

Assignment 2: A Small Numerical Library
Writeup

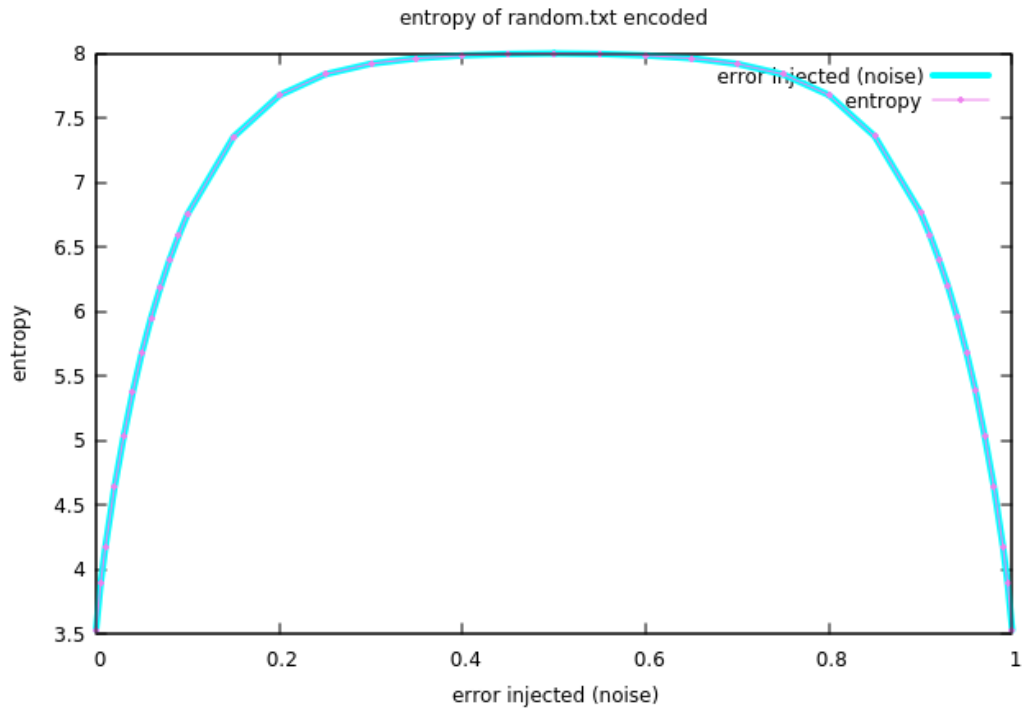
From the Assignment PDF, entropy is defined as the "amount of information that is produced by some process" (16). From the graphs in this document (with the exception of a.txt), we can see that the entropy is in a upside down parabolic shape. On top of this, the graph is fairly symmetric about the y-value of the apogee and the entropy when zero noise is injected is the same as when 100% noise is injected into the file.

Entropy Before Encoding aaa.txt: -0.000000

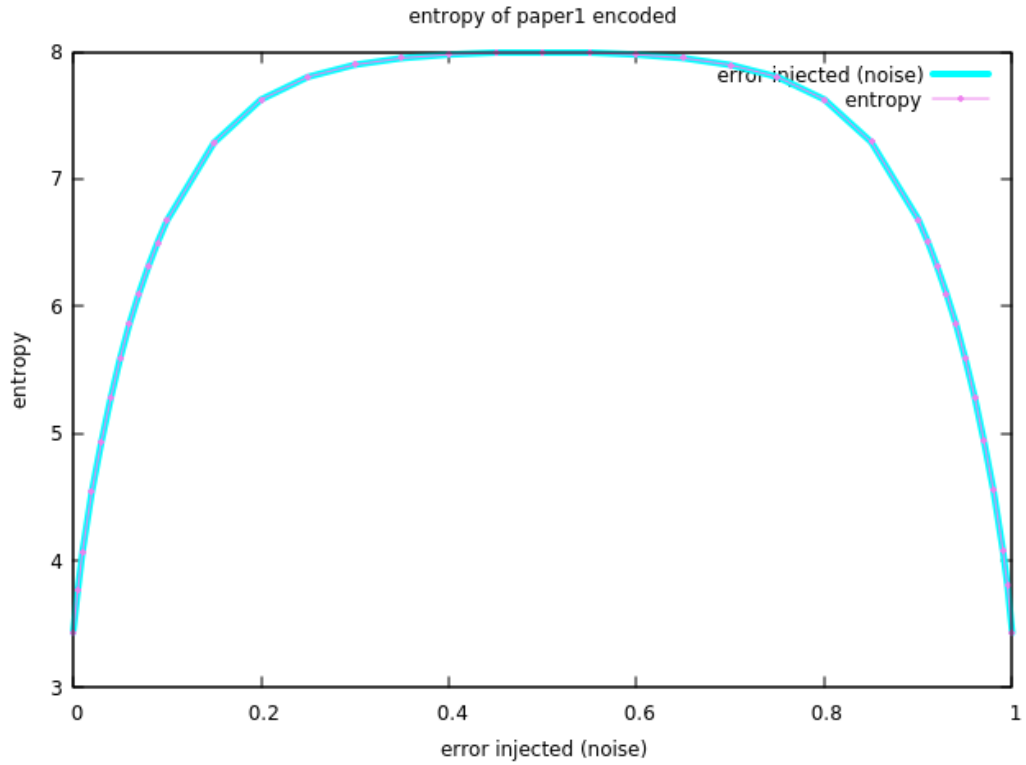


Before encoding anything in aaa.txt, we have an entropy of 0. What was different about this file when compared to the others is that this file just consisted of many of the same character, a. I suppose because of this, there wasn't any uncertainty in the information so the entropy was zero. This graph had a more round parabolic shape than the rest of the graphs which means a more linear rate of change of entropy as error is injected.

Entropy Before Encoding random.txt: 5.999488



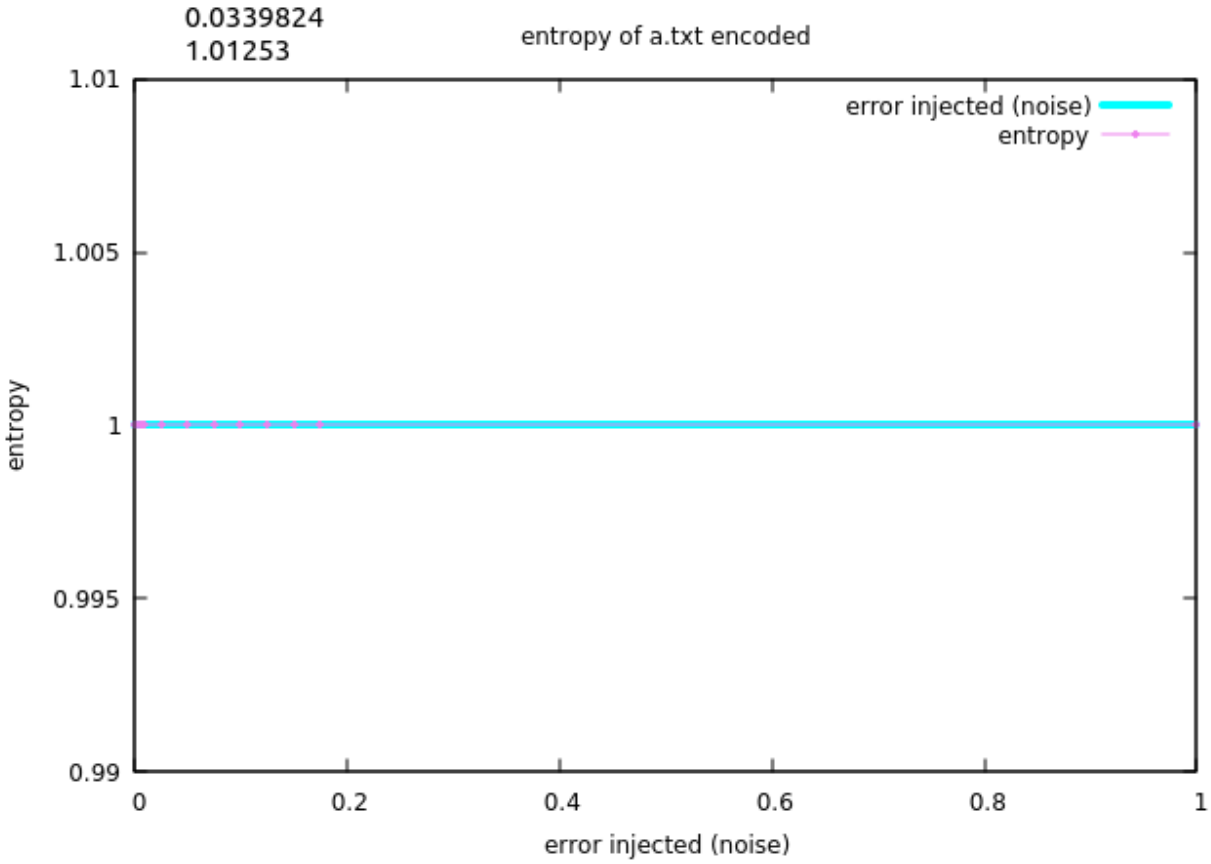
Entropy Before Encoding paper1: 4.982983



These two graphs are fairly similar to each other. They both also have a flatter top and steeper sides than the aaa.txt graph. I suppose this is due to the fact that there is more “randomness” or “variety” in these files whereas aaa.txt just had the character, a. You notice this from the entropy of these files before encoding them as well.

An explanation for why the endpoints of the graph have the same entropy and why the graph is symmetrical might be that once enough random error is injected, the errors can be interpreted as correct and seem more predictable than random.

Entropy Before Encoding a.txt: -0.000000



This file only had one character in it so before encoding, there was no level of uncertainty to the file and after encoding, there is still none because it is only one character.