### Word length Analysis

Based on comparison of three sets - Voynich and two other languages - French and Latin, there is a very similar structure. The peak occurs around 5 < length < 10, and the histogram is downward sloping after length > 10. This again supports the theory of Voynich manuscript being a real language.

### Anagrams Analysis

I have also decided to compare Voynich manuscript to English based on anagrams of length > 2. The top one is the former, and the bottom one represents the latter. Both graphs are downward sloping and again have a similar shape.

Note:

The difference of the x-axis of course amounts to the fact of the set size. I have had a much bigger size set of English from Homework #2.

### Arithmetic Compression

Most interestingly, by running arithmetic compression algorithm on Voynich manuscript in comparison to English it seems like the similarities are quite obvious. As per Homework #7, the x-axis indicate probability of starting with that phoneme, while y axis indicates probability of ending with it. In fact, we can see on the graph clear concentration of points that almost form a line in both graphs. For example, points at (x, 0.2) form almost straight perpendicular line. That means that phoneme is a common end phoneme. That makes sense for English since a lot of words have similar endings such as all adjectives with -ed at the end. However, it is quite surprising for Voynich manuscript.

**Morphological structure**

It potentially suggests a morphological structure of Voynich. Thick lines of endings and beginnings on certain letter might indicate suffixes and prefixes.

### Punctuation

From my observation on the Voynich document provided, it seems like there are two letters that occur only at the end of the sentences and that is “K” and “L”. But those also occur in the vocabulary, so there seems to be no punctuation, at least not in a common sense.

Examples:

!%%%\*!OM.OHCCG.OHCAR.ROEOHG.HZAAR.8AM.ODAM.OR.ODAL-

DTRG.TO8AM.OE.OEHTCG.TAR.FZAR.AK-

### Conclusion

Based on comparison of Voynich manuscript versus natural languages, analysis of this assignment seems to indeed indicate number of similarities from word length to arithmetic compression to the structure of anagrams.