

Introduction to Probability and Information Theory

Homework #7

1. In class, we saw a way to measure the linguistic diversity of a region using a probability model. How might you use Information Theory to do the same thing? Define an Information Theoretic measure of linguistic diversity and describe how it differs from its probabilistic counterpart.
2. Here is a (simplified) chart of noun endings in Modern Greek:

Class	Singular				Plural			
	Nom	Gen	Acc	Voc	Nom	Gen	Acc	Voc
1	-os	-u	-on	-e	-i	-on	-us	-i
2	-s	-∅	-∅	-∅	-es	-on	-es	-es
3	-∅	-s	-∅	-∅	-es	-on	-es	-es
4	-∅	-s	-∅	-∅	-is	-on	-is	-is
5	-os	-u	-o	-o	-a	-on	-a	-a
6	-∅	-u	-∅	-∅	-a	-on	-a	-a
7	-os	-us	-os	-os	-i	-on	-i	-i
8	-∅	-os	-∅	-∅	-a	-on	-a	-a

Each row gives the endings for a particular declension. You can assume that which declension a noun belongs to is completely arbitrary and that all declensions have the same probability.

- (a) What is the probability that the nom.sg form of a noun ends in *-os*?
 - (b) If the nom.sg. of a word ends in *-os*, what is the probability that its nom.pl form ends in *-i*?
 - (c) If the voc.pl of a word ends in *-a*, what is the probability that it's in declension 5?
 - (d) What is the entropy of the Acc.sg forms?
 - (e) What is the entropy of the Gen.pl. forms?
 - (f) What is the conditional entropy of the Acc.pl forms given the Gen.sg?
3. Alver Ellegård (1953) studied the occurrence of periphrastic *do* (as in 'You do walk') in affirmative declarative sentences in Middle and Early Modern English. He reports that in the period 1535--1550, there were 1,564 uses of periphrastic *do* out of a sample of 19,200 sentences. In the period 1535--1550, there were 396 uses of periphrastic *do* in a sample of 14,600 sentences. Are the frequencies of *do* in these two time periods significantly different?