3.1 RAT

We are given the following corpus, similar to the one in lecture but with "ham" replaced by "Sam" and "I am Sam" included twice:

```
<s> I am Sam </s>
<s> Sam I am </s>
<s> I am Sam </s>
<s> I do not like green eggs and Sam </s>
```

Include <s> and </s> in your counts just like any other token.

1. Using a bigram language model with add-one smoothing, what is *P*(Sam | am)?

2. Using interpolated Kneser-Ney smoothing, what is P_{KN}(Sam | am) if we use a discount factor of *d*=1?
   Here are some quantities of interest to make this less tedious:

   - *c*(am,Sam)=2
   - *c*(am)=3
   - *c*(Sam)=4
   - |{*w*:c(am,w)>0}|=2
   - |{(w_{j-1},w_j):*c*(w_{j-1},w_j)>0}|=14
   - |{w_{i-1}:*c*(w_{i-1},Sam)>0}|=3

   As a reminder, here is the formula for P_{KN}:

$$P_{KN}(w_i|w_{i-1}) = \frac{max(c(w_{i-1},w_i)-d,0)}{c(w_{i-1})} + \lambda(w_{i-1})P_{CONTINUATION}(w_i)$$

   where $$\lambda(w_{i-1}) = \frac{d}{c(w_{i-1})}|\{w : c(w_{i-1},w) > 0\}|$$

   and $$P_{CONTINUATION}(w_i) = \frac{|\{w_{i-1} : c(w_{i-1},w_i) > 0\}|}{|\{(w_{j-1},w_j) : c(w_{j-1},w_j) > 0\}|}$$

3. If we use linear interpolation smoothing between a maximum-likelihood bigram model and a maximum-likelihood unigram model with $\lambda_1=½$ and $\lambda_2=½$, what is P(Sam|am)? Include <s> and </s> in your counts just like any other token.

4. Suppose we train a trigram language model with add-one smoothing on a given corpus. The corpus contains V word types. What is $P(w_3|w_1,w_2)$, where w_3 is a word which follows the bigram (w_1,w_2)? We use the notation $c(w_1,w_2,w_3)$ to denote the number of times that trigram (w_1,w_2,w_3) occurs in the corpus, and so on for bigrams and unigrams.