N_Grams
Week 2, Natural Language Processing
Jonathan Stewart

1)

```
<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>
```

The formula for a Tri-gram model is:

$$P(W_n|W_{n-1}, W_{n-2}) = C(W_i, W_{i-1}, W_{i-2})/C(W_{i-1}, W_{i-2})$$

All non-zero tri-gram probabilities include:

$P(< s > |Sam, < /s >) = 2/3$
$P(am|I, < s >) = 2/3$
$P(Sam|am, I) = 2/3$
$P(< /s > |Sam, am) = 2/2$
$P(Sam| < s >, < /s >) = 1/3$
$P(I|Same, < s >) = 1/1$
$P(am|I, Sam) = 1/1$
$P(< /s > |am, I) = 1/1$
$P(I| < s >, < /s >) = 2/3$
$P(do|I, < s >) = 1/3$
$P(not|do, I) = 1/1$
$P(like|not, do) = 1/1$
$P(green|like, not) = 1/1$
$P(eggs|green, like) = 1/1$
$P(and|eggs, green) = 1/1$
$P(Sam|and, eggs) = 1/1$
$P(< /s > |Sam, and) = 1/1$

2)
$(< s >)(i)(want)(to)(eat)(chinese)(food) < /s >$
$(.25)(.33)(.66)(.28)(.021)(.52)(.68) = .0001132$

Smoothed:

$$(<s>)(i)(want)(to)(eat)(chinese)(food)</s>$$
$$(.235)(.21)(.26)(.18)(.0078)(.052)(.62) = .0000005807$$

3)
The non-smoothed probability was much higher. This is because the smoothed version acts to make all zero-probability occurrences of words in a given vocabulary non-zero, which dissipates probability away from non-zero steps.Because markov chains are row normalized to always equal 1, increasing a probability for a given step necessitates decreasing it for another. Since the sentence, ‘I want to eat chinese food’ was a combination of words that had all occurred in sequence for a 1-gram model, The probability of that overall sequence was decreased in favor of all sequences for which zero probabilities occurred.

4)

$$V = 11$$
$$P(Sam|am) = (2+1)/(3+11) = .214$$