Problem 1: Laplace Smoothing

Let $|V| = 5$ where V is the vocabulary and $N = \sum_{i}^{|V|} freq(w_i) = 1,200$ where N is the number of words in the corpus. Without smoothing, the probabilities of each noun in the corpus are:

$P(maple) = \frac{freq(maple)}{N} = \frac{600}{1{,}200} = 0.5$

$P(oak) = \frac{freq(oak)}{N} = \frac{400}{1{,}200} = 0.333$

$P(pine) = \frac{freq(pine)}{N} = \frac{180}{1{,}200} = 0.15$

$P(spruce) = \frac{freq(spruce)}{N} = \frac{20}{1{,}200} = 0.017$

$P(aspen) = \frac{freq(aspen)}{N} = \frac{0}{1{,}200} = 0$

Using Laplace smoothing (aka: Add-One Smoothing), the new frequencies for each noun:

$new\_freq(maple) = (freq(maple) + 1) * \frac{N}{N+V} = (600 + 1) * \frac{1{,}200}{1{,}200+5} = \frac{721{,}200}{1{,}205} = 598.506$

$new\_freq(oak) = (freq(oak) + 1) * \frac{N}{N+V} = (400 + 1) * \frac{1{,}200}{1{,}200+5} = \frac{481{,}200}{1{,}205} = 399.336$

$new\_freq(pine) = (freq(pine) + 1) * \frac{N}{N+V} = (180 + 1) * \frac{1{,}200}{1{,}200+5} = \frac{217{,}200}{1{,}205} = 180.249$

$new\_freq(spruce) = (freq(spruce) + 1) * \frac{N}{N+V} = (20 + 1) * \frac{1{,}200}{1{,}200+5} = \frac{25{,}200}{1{,}205} = 20.913$

$new\_freq(aspen) = (freq(aspen) + 1) * \frac{N}{N+V} = (0 + 1) * \frac{1{,}200}{1{,}200+5} = \frac{1{,}200}{1{,}205} = 0.966$

Given these new frequencies, the new probabilities are calculated:

$new\_P(maple) = \frac{new\_freq(maple)}{N} = \frac{598.506}{1{,}200} = 0.499$

$new\_P(oak) = \frac{new\_freq(oak)}{N} = \frac{399.336}{1{,}200} = 0.333$

$new\_P(pine) = \frac{new\_freq(pine)}{N} = \frac{180.249}{1{,}200} = 0.15$

$new\_P(spruce) = \frac{new\_freq(spruce)}{N} = \frac{20.913}{1{,}200} = 0.017$

$new\_P(aspen) = \frac{new\_freq(aspen)}{N} = \frac{0.996}{1{,}200} = 0.001$

Completing the table:

Noun	**Freq.**	**Unsmoothed Prob.**	**Smoothed Freq.**	**Smoothed Prob.**
maple	600	$\frac{600}{1{,}200} = 0.5$	$\frac{721{,}200}{1{,}205} = 598.506$	$\frac{598.506}{1{,}200} = 0.499$
oak	400	$\frac{400}{1{,}200} = 0.333$	$\frac{481{,}200}{1{,}205} = 399.336$	$\frac{399.336}{1{,}200} = 0.333$
pine	180	$\frac{180}{1{,}200} = 0.15$	$\frac{217{,}200}{1{,}205} = 180.249$	$\frac{180.249}{1{,}200} = 0.15$
spruce	20	$\frac{20}{1{,}200} = 0.017$	$\frac{25{,}200}{1{,}205} = 20.913$	$\frac{20.913}{1{,}200} = 0.017$
aspen	0	$\frac{0}{1{,}200} = 0$	$\frac{1{,}200}{1{,}205} = 0.996$	$\frac{0.996}{1{,}200} = 0.001$

Table 1: Unsmoothed and smoothed probabilities and frequencies.

2: Grammars and Recursive Transition Networks

(a) Grammar A and Grammar B - **DIFFERENT**

Grammar A requires the NP to begin with an article but Grammar B does not.

An example POS tag sequence **accepted by Grammar B** and not by Grammar A is: **noun**.

(b) Grammar A and Grammar C - **DIFFERENT**

Grammar C requires one or more adjectives after the article but Grammar A requires zero or more adjectives after the article.

An example POS tag sequence **accepted by Grammar A** and not by Grammar C is: **art noun**.

(c) Grammar A and RTN-2 - **DIFFERENT**

Grammar A requires the NP to begin with an article but RTN-2 does not.

An example POS tag sequence **accepted by RTN-2** and not by Grammar A is: **noun**.

(d) Grammar A and RTN-3 - **DIFFERENT**

RTN-3 requires one or more adjectives after the article but Grammar A requires zero or more adjectives after the article.

An example POS tag sequence **accepted by Grammar A** and not by RTN-3 is: **art noun**.

(e) Grammar B and RTN-2 - **SAME**

(f) Grammar C and RTN-1 - **DIFFERENT**

Grammar C requires the NP to end with one or more nouns but RTN-1 does not.

An example POS tag sequence **accepted by RTN-1** and not by Grammar C is: **art adj**.

(g) Grammar C and RTN-3 - **SAME**

(h) RTN-1 and RTN-3 - **DIFFERENT**

RTN-3 requires the NP to end with one or more nouns but RTN-1 does not.

An example POS tag sequence **accepted by RTN-1** and not by Grammar C is: **art adj**.

3: N-Gram Probabilities

Let k denote the number of distinct lexical unigrams in the text corpus.

(a) $P(the) = \frac{freq(the)}{\sum_i^k freq(w_i)} = \boxed{\frac{5}{34}}$

Let k denote the number of distinct POS unigrams in the text corpus.

(b) $P(the) = \frac{freq(VERB)}{\sum_i^k freq(T_i)} = \boxed{\frac{6}{34}}$

Let $freq(W_{n-m}, ..., W_{n-1}, W_n)$ denote the frequency of the phrase $W_{n-m}, ..., W_{n-1}, W_n$ appearing in the text corpus.

(c) $P(young \mid girl) = \frac{freq(W_{n-1}, W_n)}{freq(W_{n-1})} = \frac{freq(girl\ young)}{freq(girl)} = \boxed{\frac{0}{3}}$

(d) $P(girl \mid young) = \frac{freq(W_{n-1}, W_n)}{freq(W_{n-1})} = \frac{freq(young\ girl)}{freq(young)} = \boxed{\frac{2}{2}}$

(e) $P(and \mid women) = \frac{freq(W_{n-1}, W_n)}{freq(W_{n-1})} = \frac{freq(women\ and)}{freq(women)} = \boxed{\frac{1}{3}}$

(f) $P(thanked \mid young\ girl) = \frac{freq(W_{n-2}, W_{n-1}, W_n)}{freq(W_{n-2}, W_{n-1})} = \frac{freq(young\ girl\ thanked)}{freq(young\ girl)} = \boxed{\frac{0}{2}}$

(g) $P(five \mid gave\ her) = \frac{freq(W_{n-2}, W_{n-1}, W_n)}{freq(W_{n-2}, W_{n-1})} = \frac{freq(gave\ her\ five)}{freq(gave\ her)} = \boxed{\frac{1}{2}}$

Let $C(event)$ denote the count of the given event occurring in the text corpus.

(h) $P(the \mid ART) = \frac{C(w_i\ with\ tag\ T_i)}{C(any\ word\ with\ tag\ T_i)} = \frac{C(the\ with\ tag\ ART)}{C(any\ word\ with\ tag\ ART)} = \boxed{\frac{5}{8}}$

(i) $P(cross \mid NOUN) = \frac{C(w_i\ with\ tag\ T_i)}{C(any\ word\ with\ tag\ T_i)} = \frac{C(cross\ with\ tag\ NOUN)}{C(any\ word\ with\ tag\ NOUN)} = \boxed{\frac{0}{9}}$

(j) $P(thanked \mid VERB) = \frac{C(thanked\ with\ tag\ VERB)}{C(any\ word\ with\ tag\ T_i)} = \frac{C(thanked\ with\ tag\ VERB)}{C(any\ word\ with\ tag\ VERB)} = \boxed{\frac{2}{6}}$

(k) $P(NUM \mid PRO) = \frac{C(T_i\ immediately\ follows\ T_{i-1})}{C(any\ tag\ immediately\ follows\ T_{i-1})} = \frac{C(NUM\ immediately\ follows\ PRO)}{C(any\ tag\ immediately\ follows\ PRO)} = \boxed{\frac{1}{2}}$

(l) $P(ART \mid VERB) = \frac{C(T_i\ immediately\ follows\ T_{i-1})}{C(any\ tag\ immediately\ follows\ T_{i-1})} = \frac{C(ART\ immediately\ follows\ VERB)}{C(any\ tag\ immediately\ follows\ VERB)} = \boxed{\frac{4}{6}}$

4: Viterbi Algorithm

(a) $P(light = VERB) = P(VERB \mid \phi) * P(light \mid VERB) = 0.25 * 0.5 = \boxed{0.125}$

(b) $P(light = NOUN) = P(NOUN \mid \phi) * P(light \mid NOUN) = 0.7 * 0.6 = \boxed{0.42}$

(c) $P(light = ADJ) = P(ADJ \mid \phi) * P(light \mid ADJ) = 0.15 * 0.2 = \boxed{0.03}$

(d) $P(shows = VERB) = P(shows \mid V) * MAX[(P(light = V) * P(V \mid V)), (P(light = N) * P(V \mid N)), (P(light = A) * P(V \mid A))]$

$P(shows = VERB) = 0.3 * MAX[(0.125 * 0.4), (0.42 * 0.5), (0.03, 0.1)] = \boxed{0.063}$

(e) $P(shows = NOUN) = P(shows \mid N) * MAX[(P(light = V) * P(V \mid V)), (P(light = N) * P(V \mid N)), (P(light = A) * P(VE \mid A))]$

$P(shows = NOUN) = 0.4 * MAX[(0.125 * 0.3), (0.42 * 0.8), (0.03, 0.6)] = \boxed{0.1344}$

(f) $P(shows = ADJ) = P(shows \mid A) * MAX[(P(light = V) * P(V \mid V)), (P(light = N) * P(V \mid N)), (P(light = A) * P(V \mid A))]$

$P(shows = ADJ) = 0.1 * MAX[(0.125 * 0.7), (0.42 * 0.2), (0.03, 0.9)] = \boxed{0.00875}$

The most likely sequence of POS tags for the sentence *"Light shows"* is Light=NOUN and shows=NOUN.

light=NOUN and shows=NOUN