

B. Problems.

1. (a) Regular Expression:

$$(a(b(ab)))^+$$

(b) Regular Expression:

$$(ab a?)^+$$

2. Bigram Probabilities:

(a) $\langle s \rangle$ a man a man a man a plan a plan
a canal panama panama $\langle /s \rangle$

Unigram counts:

$\langle s \rangle$	a	man	plan	canal	panama	$\langle /s \rangle$
1	6	3	2	1	2	1

Bigram counts:

$$c(\langle s \rangle a) = 1 \quad c(a \text{ man}) = 3$$

$$c(\text{man } a) = 3 \quad c(a \text{ plan}) = 2$$

$$c(\text{plan } a) = 2 \quad c(a \text{ canal}) = 1$$

$$c(\text{canal panama}) = 1 \quad c(\text{panama panama}) = 1$$

$$c(\text{panama } \langle /s \rangle) = 1$$

(b) $\langle s \rangle$ plan a panama $\langle /s \rangle$

(i) No smoothing.

$$\begin{aligned}
 P(\text{plan } a \text{ panama}) &= P(\text{plan} | \langle s \rangle) P(a | \text{plan}) \\
 &\quad P(\text{panama} | a) P(\langle /s \rangle | \text{panama}) \\
 &= \frac{0}{1} * \frac{2}{2} * \frac{0}{6} * \frac{1}{2} = 0
 \end{aligned}$$

(ii) Add one smoothing

$$c_i^* = (c_i + 1) \frac{N}{N+V} \quad V = 7$$

$$\begin{aligned} P(\text{plan a panama}) &= \left(\frac{c(\langle s \rangle \text{ plan}) + 1}{c(\langle s \rangle) + V} \right) \\ &\quad \left(\frac{c(\text{plan a}) + 1}{c(\text{plan}) + V} \right) \left(\frac{c(\text{a panama}) + 1}{c(\text{a}) + V} \right) \\ &\quad \left(\frac{c(\text{panama } \langle /s \rangle) + 1}{c(\text{panama}) + V} \right) \\ &= \left(\frac{0+1}{1+7} \right) \left(\frac{2+1}{2+7} \right) \left(\frac{0+1}{6+7} \right) \left(\frac{1+1}{2+7} \right) \\ &= 0.00071 \end{aligned}$$

(iii) Good Turing discounting based smoothing

$$c = 0 \quad N_0 = 40 = (7 \times 7) - (5 + 2 + 2)$$

$$c = 1 \quad N_1 = 5$$

$$c = 2 \quad N_2 = 2$$

$$c = 3 \quad N_3 = 2$$

$$N = 5 + (2 \times 2) + (3 \times 2) = 15$$

$$P(\langle s \rangle \text{ plan}) \Rightarrow c = 0 \quad p = \frac{N_1}{N} = \frac{5}{15}$$

$$p(\text{plan a}) \Rightarrow c = 2 \quad p = \frac{(c+1)N_{c+1}}{N_c \times N} = \frac{3}{15}$$

$$p(\text{a panama}) \Rightarrow c = 0 \quad p = \frac{N_1}{N} = \frac{5}{15}$$

$$p(\text{panama } \langle /s \rangle) \Rightarrow c = 1 \quad p = \frac{4/5}{15}$$

$$P(\text{plan a panama}) = \left(\frac{5}{15} \right) \left(\frac{3}{15} \right) \left(\frac{5}{15} \right) \left(\frac{0.8}{15} \right) = 0.00118$$