

①

4) mississippi shipping

letters	count	Est. Frequency
space	1	$1/20 = 0.05$
g	1	$1/20 = 0.05$
h	1	$1/20 = 0.05$
i	6	$6/20 = 0.3$
m	1	$1/20 = 0.05$
n	1	$1/20 = 0.05$
p	4	$4/20 = 0.2$
s	5	$5/20 = 0.25$
other	0	$0/20 = 0$
total: 20		1

$$P(w_i) = (c(w_i)/N)$$

$$p(w) = 1/20 \cdot 1/20 \cdot 1/20 \cdot 6/20 \cdot 1/20 \cdot 1/20 \cdot 4/20 \cdot 5/20 \cdot 0/20$$

$$\underline{P(w) = 0}$$

b) Now with Laplace add-one smoothing

letters	count + 1	Est. Freq.
space	2	$2/47 = 0.043$
g	2	$2/47 = 0.043$
h	2	$2/47 = 0.043$
i	7	$7/47 = 0.15$
m	2	$2/47 = 0.043$
n	2	$2/47 = 0.043$
p	5	$5/47 = 0.106$
s	6	$6/47 = 0.127$
others	1.19	$19/47 = 0.404$

$$P(w) = (C(w) + 1) / (N + V)$$

V = number of unique characters

$$= \frac{2}{47} \cdot \frac{2}{47} \cdot \frac{2}{47} \cdot \frac{7}{47} \cdot \frac{2}{47} \cdot \frac{2}{47} \cdot \frac{5}{47} \cdot \frac{6}{47} \cdot \frac{19}{47}$$

V = total no of char. in training sample

$$= 127680 / 1.12 \times 10^{15}$$

(c)

(letter) characters	count	Est. freq
space	1/28	0.03571
g	1/28	0.03571
h	1/28	0.03571
i	6/28	0.21428
m	1/28	0.03571
n	1/28	0.03571
p	4/28	0.14285
s	5/28	0.17857
new letters	8/28	0.28571
total	28	

char.	Estid. freq.
space	0.03571
g	0.03571
h	0.03571
i	0.21428
m	0.03571
n	0.03571
p	0.14285
s	0.17857
other letters	0.01503 ($8/28 \cdot 19$)

$$\text{other letters} = \frac{8}{28 \cdot (27-8)}$$

$$= \frac{8}{28 \cdot 19}$$

$$\hookrightarrow p(w) = \frac{1}{28} \cdot \frac{1}{28} \cdot \frac{1}{28} \cdot \frac{6}{28} \cdot \frac{1}{28} \cdot \frac{4}{28} \cdot \frac{5}{28} \cdot \frac{8}{532}$$

$$\approx \underline{4.7706616 \cdot 10^{-13}}$$