A3Q4

Note *
representing space character as ' '

phrase = 'to be or not to be'

alphabet = 26
space char = 1 } #|V| = 27

1 No smoothing probabilities

x	#X	P(x)
t	III = 3	3/18
o	IIII = 4	4/18
b	II = 2	2/18
e	II = 2	2/18
r	I = 1	1/18
n	I = 1	1/18
' '	IIIII = 5	5/18
*	0 = 0	0

space character ←

anything else

(2) propabilities after caplace smoothing

caplace smoothing is to add '1' to all events in beginning.

X	#(x)		P(x)
t	1111	= 4	4/45
o	11111	= 5	5/45
b	111	= 3	3/45
c	111	= 3	3/45
r	11	= 2	2/45
n	11	= 2	2/45
' '	111111	= 6	6/45
a	1	= 1	1/45
b	1	= 1	1/45
c	1	= 1	1/45
⋮	⋮	⋮	⋮
z	1	= 1	1/45
	27 + 18	= 45	

③ Witten-Bell smoothing

so, let us say we start traund this sequence and for every new character we mark it.

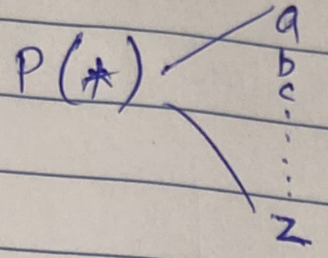
t o ' ' b e o r ' ' n o t ' ' t o b e

t o*' '*b*c ' ' o*r ' '*n o t ' ' t o ' ' bc

so whenever we see new characters we mark those.

X	# X		P(x)
*	|||||||	= 7	7/25
t	|||	= 3	3/25
o	||||	= 4	4/25
b	||	= 2	2/25
c	||	= 2	2/25
r	|	= 1	1/25
n	|	= 1	1/25
' '	|||||	= 5	5/25
	25		

$P(*) = 7/25$ } So, this is probability of all unseen events



$= 7/25$

So, probability of each unseen event will be

$$\frac{7/25}{(27-7)} = \frac{7/25}{20} = \frac{7}{25\times20}$$

$$\boxed{P(\text{unseen}) = \frac{7}{500}}$$

probability of single unseen event.