CS4248 Assignment 1

Heng Low Wee (U096901R)

1.

$$p(k) = \frac{\binom{n}{k}\binom{N-n}{n-k}}{\binom{N}{n}}$$

2.

$$P_{wb}(w|c_i) = \frac{C(c_i, w)}{C(c_i) + T(c_i)} \text{ if } C(c_i, w) > 0$$
$$P_{wb}(w|c_i) = \frac{T(c_i)}{Z(c_i) \cdot (C(c_i) + T(c_i))} \text{ if } C(c_i, w) = 0$$

So, we need to know $C(c_i, w), C(c_i), T(c_i)$ & $Z(c_i)$
$V = \{$John, loves, swimming, strengthens, our, body, jogging, is, fun, Mary$\}$
$|V| = 10$
c_1 :
$C(c_1) = |c_1| = 7$
$T(c_1) = |\forall w \in V, w \in c_1| = 6$
$Z(c_1) = |\forall w \in V, w \notin c_1| = 4$
c_2 :
$C(c_2) = |c_2| = 6$
$T(c_2) = |\forall w \in V, w \in c_2| = 5$
$Z(c_2) = |\forall w \in V, w \notin c_2| = 5$

	$P_{wb}(w\|c_1)$		$P_{wb}(w\|c_2)$	
	$C(c_1, w)$	wb-Smooth	$C(c_2, w)$	wb-Smooth
body	1	$\frac{1}{7+6} = 0.0769$	0	$\frac{5}{5\times(6+5)} = 0.0909$
fun	0	$\frac{6}{4\times(7+6)} = 0.115$	1	$\frac{1}{6+5} = 0.0909$
is	0	0.115	1	0.0909
jogging	0	0.115	2	0.182
John	1	0.0769	0	0.0909
loves	1	0.0769	1	0.0909
Mary	0	0.115	1	0.0909
our	1	0.0769	0	0.0909
strengthens	1	0.0769	0	0.0909
swimming	2	0.154	0	0.0909

3. Minimum Edit Distance

p	5	4	3	4	3	4
a	4	3	2	3	4	5
e	3	2	1	2	3	4
h	2	1	2	3	4	5
c	1	2	3	4	5	6
X	0	1	2	3	4	5
X	X	h	e	l	p	s

4. Given:
   $H(X) = -\sum_{x \in X} p(x) \log p(x)$
   $H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$
   $H(Y|X) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y|x)$
   To verify:
   $H(X,Y) = H(X) + H(Y|X)$

$$
\begin{aligned}
H(X,Y) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y) \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left[p(y|x)p(x)\right] \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y)(\log p(y|x) + \log p(x)) \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y|x) + p(x,y) \log p(x) \\
&= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y|x) + \left(-\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x)\right) \\
&= \left(-\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x)\right) + \left(-\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y|x)\right) \\
&= \left(-\sum_{x \in X} p(x) \log p(x)\right) + \left(-\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(y|x)\right) \\
&= H(X) + H(Y|X)
\end{aligned}
$$