CS4248Assignment 1

U096883L Shawn Tan

1.

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

2. Some Witten-Bell smoothing shit.

	$P_{WB}(w\|c_1)$	$P_{WB}(w\|c_2)$
body		
fun		
is		
jogging		
John		
loves		
Mary		
our		
strengthens		
swimming		

3. Table for 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
	0	1	2	3	4	5
		h	e	l	p	s

4.

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

Marginalising over y,

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