MATH 4175 - Homework 6

John Murphy
15th October 2014

1 Part A

Given a system with the following plain text distribution, we must find the entropy.

\[ P(”a”) = 0.5 \]
\[ P(”b”) = 0.25 \]
\[ P(”c”) = 0.25 \]

The shift cipher is applied to letter pairs. Each letter is independent of each other, so we can find the probability of each letter pair by multiplying probabilities together.

\[ P(”aa”) = 0.25 \]
\[ P(”ab”) = P(”ac”) = P(”ba”) = P(”ca”) = 0.125 \]
\[ P(”bb”) = P(”cb”) = P(”bc”) = P(”cc”) = 0.0625 \]

We can use these probabilities to find the entropy of the plaintext.

\[ H(\theta) = - \sum_x P(x) \log_2 P(x) \]

\[ H(\theta) = -[(0.25)\log_2(0.25) + 4 \cdot (0.125)\log_2(0.125) + 4 \cdot (0.0625)\log_2(0.0625)] \]

\[ H(\theta) = 3.0 \]
2 Part B

Now we must find \( H(P|C) \). Let’s start by looking at the probability of \( P(C=“aa”) \). That is, the probability of the cipher text equaling “aa”.

If the cipher text is “aa”, then there are three possibilities for plain text. The plain text can either be “aa”, “bb”, or “cc”. The probability of each of these digrams appearing in plaintext is \( \frac{1}{4} \) for “aa”, \( \frac{1}{16} \) for “bb” and \( \frac{1}{16} \) for “cc”.

Each of these plaintext digrams have a \( \frac{1}{3} \) chance of matching the cipher text “aa”. Using this, we can find the joint probability for each event.

\[
P(C = "aa", P = "aa") = \frac{1}{4} + \frac{1}{3} = \frac{1}{12} \tag{6}
\]

\[
P(C = "aa", P = "bb") = \frac{1}{16} + \frac{1}{3} = \frac{1}{48} \tag{7}
\]

\[
P(C = "aa", P = "bc") = \frac{1}{16} + \frac{1}{3} = \frac{1}{48} \tag{8}
\]

If we sum these probabilities, we can find the value for \( P(C = "aa") \).

\[
P(C = "aa") = \frac{1}{8} \tag{9}
\]

And finally, using this, we can find the conditional probability:

\[
P(P = "aa"|C = "aa") = \frac{2}{3} \tag{10}
\]

\[
P(P = "bb"|C = "aa") = \frac{1}{6} \tag{11}
\]

\[
P(P = "cc"|C = "aa") = \frac{1}{6} \tag{12}
\]

Note that this probability distribution is exactly the same for cipher texts “bb” and “cc”.

Now let us examine the case where the cipher text is equal to “ab”. We can follow the same process as before to find conditional probabilities.

\[
P(C = "ab", P = "ab") = \frac{1}{8} + \frac{1}{3} = \frac{1}{24} \tag{13}
\]

\[
P(C = "ab", P = "bc") = \frac{1}{16} + \frac{1}{3} = \frac{1}{48} \tag{14}
\]

\[
P(C = "ab", P = "ca") = \frac{1}{8} + \frac{1}{3} = \frac{1}{24} \tag{15}
\]
We can no sum these probabilities to find the $P(C = "ab")$.

$$P(C = "ab") = \frac{5}{48} \quad \text{(16)}$$

And now we can use this sum to find conditional probabilities.

$$P(P = "ab"|C = "ab") = \frac{2}{5} \quad \text{(17)}$$

$$P(P = "bc"|C = "ab") = \frac{1}{5} \quad \text{(18)}$$

$$P(P = "ca"|C = "ab") = \frac{2}{5} \quad \text{(19)}$$

For the sake of brevity, note that this distribution is the same for cipher-texts “bc”, “ca”, “ac”, “ba”, and “cb”.

Now that we have the joint and conditional probabilities of each possible cipher text and plain text, we can use the entropy equation below to find the entropy $H(P|C)$.

$$H(Y|X) = \sum_{x,y} P(x,y)\log_2 P(y|x) \quad \text{(20)}$$

Substituting in the above values, we find the following.

$$H(P|C) = 3[\frac{1}{12}\log_2(\frac{2}{3}) + 2\frac{1}{48}\log_2(\frac{1}{6})] + 6[2\frac{1}{24}\log_2(\frac{2}{5}) + \frac{1}{48}\log_2(\frac{1}{5})] \quad \text{(21)}$$

$$H(P|C) = 1.421 \quad \text{(22)}$$

### 3 Part C

In part A, we found that the entropy of the plaintext was $H(P) = 3.0$. In part B, we found the conditional entropy, $H(P|C) = 1.421$.

By comparing the two entropies, we can see that there our shift cipher revealed a large amount of information about the plaintext. About 1.5 “bits” of information were revealed.

This suggests that our crypto-system revealed slightly over half of the information that was encrypted.