THE UNIVERSITY OF SYDNEY
MATH3024 ELEMENTARY CRYPTOGRAPHY AND PROTOCOLS

Semester 1	**Exercises for Week 4**	2004

Information Theory

In this tutorial we consider the information theory of languages. In order to understand naturally occurring languages, we consider the models for finite languages $\mathcal{L}$ consisting of strings of fixed finite length N together with a probability function P which models the natural language. In what follows, for two strings X and Y we denote their concatenation by XY.

1. Consider the language of 1-character strings over $\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D}\}$ with associated probabilities 1/3, 1/12, 1/4, and 1/3. What is its corresponding entropy?

2. Consider the language $\mathcal{L}_2$ of all strings of length 2 in $\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D}\}$ defined by the probability function of Exercise 1 and 2-character independence: $P(XY) = P(X)P(Y)$. What is the entropy of this language?

3. Let $\mathcal{M}$ be the strings of length 2 over $\{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D}\}$ with the following frequency distribution:

$$\begin{array}{llll}
P(\mathtt{AA}) = 5/36 & P(\mathtt{BA}) = 0 & P(\mathtt{CA}) = 1/12 & P(\mathtt{DA}) = 1/9 \\
P(\mathtt{AB}) = 1/36 & P(\mathtt{BB}) = 1/144 & P(\mathtt{CB}) = 1/48 & P(\mathtt{DB}) = 1/36 \\
P(\mathtt{AC}) = 7/72 & P(\mathtt{BC}) = 1/48 & P(\mathtt{CC}) = 1/16 & P(\mathtt{DC}) = 5/72 \\
P(\mathtt{AD}) = 5/72 & P(\mathtt{BD}) = 1/18 & P(\mathtt{CD}) = 1/12 & P(\mathtt{DD}) = 1/8
\end{array}$$

Show that the 1-character frequencies in this language are the same as for the language in Exercise 2.

4. Do you expect the entropy of the language of Exercise 3 to be greater or less than that of Exercise 2? What is the entropy of each language?

5. Consider the infinite language of all strings over the alphabet $\{\mathtt{A}\}$, with probability function defined such that $P(\mathtt{A}\dots\mathtt{A}) = 1/2^n$, where n is the length of the string $\mathtt{A}\dots\mathtt{A}$. Show that the entropy of this language is 2.

Frequency Analysis

Consider those ciphertexts from the last tutorial which come from a Vigènere cipher. Use the javascript application for analyzing Vigenère ciphers:

```
http://magma.maths.usyd.edu.au/~kohel/
                    teaching/MATH3024/Javascript/vigenere.html
```

to determine the periods and keys for each of the ciphertext samples.