Do 5 problems

1. Do one of (a) or (b). In either case, give enough explanation so that I know how you arrived at your solution.

(a) The following English-language message was encrypted using a simple substitution cipher. Decipher the message. You can also obtain the ciphertext as the file hw2-1.txt in the homework directory.

```
EGOWZCOWECELCWYDYQWYZHCMPZYHKCYEQDPWTWPYBWCMPHZMTG
HMZPWHYBGCJWEYXEGOWZPCECEGMECHMZPGYQWQGTwZPYPHGHWH
CMPZJPyTLWEQEPILS
```

(b) Decipher the following “message”. The plaintext is an ordinary long-division computation in base 10, which has been encrypted using a simple substitution cipher. In this case, the digits 0 through 9 have been replaced by letters of the alphabet. Present your answer by giving the encryption of the digits 0, 1, …, 9. Note: no computers allowed for this problem.

```
SEY
EAT
YPOMC
YEC
RRM
CMY
MSRC
MECT
MAY
```

2. I have encrypted three passages of English-language text. One is encrypted using a simple substitution cipher, one with a polyalphabetic cipher and one with a permutation cipher. Your job is to figure out which one is which. Give a short explanation of the basis for your decision.

The text is available in the homework directory as hw2-2.txt. For your convenience I have precomputed the frequency distribution of the individual characters in each piece of ciphertext:

```
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>text 1</td>
<td>27</td>
<td>6</td>
<td>11</td>
<td>8</td>
<td>44</td>
<td>11</td>
<td>9</td>
<td>18</td>
<td>34</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>text 2</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>34</td>
<td>24</td>
<td>13</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>text 3</td>
<td>22</td>
<td>7</td>
<td>13</td>
<td>11</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td>12</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>N</td>
<td>O</td>
<td>P</td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
<td>U</td>
<td>V</td>
<td>W</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>text 1</td>
<td>34</td>
<td>39</td>
<td>8</td>
<td>0</td>
<td>29</td>
<td>28</td>
<td>34</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>text 2</td>
<td>2</td>
<td>6</td>
<td>20</td>
<td>5</td>
<td>4</td>
<td>22</td>
<td>25</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>text 3</td>
<td>6</td>
<td>14</td>
<td>17</td>
<td>22</td>
<td>13</td>
<td>10</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>16</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>
```
3. I have taken a chunk of text from a newspaper and encrypted it with a polyalphabetic cipher. Use Kasiski’s method to suggest the blocksize. Then compute the index of coincidence to double-check the answer you got from Kasiski. You can get the text from the homework directory as hw2-3.txt. You may use a computer to do the computations. Note: you do not have to decrypt the passage.

4. Let us define the following encryption scheme. Take $\mathcal{M} = \mathcal{C} = \{0, 1, 2, 3\}$ and $\mathcal{K} = \{1, 2, 3\}$. The probability distributions for the plaintexts and the keys are

<table>
<thead>
<tr>
<th></th>
<th>$P(M)$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>$\frac{1}{8}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{3}{8}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$P(K)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td></td>
</tr>
</tbody>
</table>

The encryption of plaintext $i$ by key $j$ is equal to $(i + j) \mod 4$. For this problem, we are only encrypting a single character, not strings of characters.

(a) Explain why the values of $\mathcal{M}$ and $\mathcal{C}$ determine $\mathcal{K}$. Using this fact, argue that $H(K|M, C) = 0$, and show that $H(M|C) = H(K|C)$.

(b) Compute the values of $H(M)$, $H(C)$, $H(K)$, and $H(K|C)$.

(c) Is this cryptosystem perfect?

5. Suppose that English text is encrypted with a permutation cipher of block size $d$ (and assume that $d$ is “large”). As usual, $C^n$ denotes a random $n$-gram of ciphertext.

(a) What do you think will be the value of $H(C^1)$? Justify your answer. What about $H(C^n)$ for $n > 1$.

(b) Find a formula for the unicity point in terms of $d$. Compute for $d = 5, 10, 15, 20$.

6. The purpose of this exercise is to prove the assertion I made in class (Unit 11.3): Let $M^n = \langle M_1, M_2, \ldots, M_n \rangle$ denote a typical $n$-gram of English language text. Assume that $\lim_{n \to \infty} H(M_n|M_1, M_2, \ldots, M_{n-1})$ exists, and is equal to $L$. Then $\lim_{n \to \infty} H(M^n)/n$ exists, and is also equal to $L$.

(a) First prove the following lemma: For any sequence $a_1, a_2, \ldots$ of real numbers,

$$\lim_{n \to \infty} a_n = a \implies \lim_{n \to \infty} \frac{a_1 + a_2 + \cdots + a_n}{n} = a.$$  

(If you can’t prove the lemma, you may still use it to do the rest of the problem.)

(b) Let $u_n = H(M_n|M_1, \ldots, M_{n-1})$ and $h_n = H(M^n)$. Then $h_n = h_{n-1} + u_n$, hence $h_n = u_1 + u_2 + \cdots + u_n$.

(c) Prove the theorem.