Solutions 2010

1

(a) The most frequently occurring letter is 'X'. Since the most frequently occurring letter in English is 'E', we shall begin by assuming that

Cleartext 'E' $\longmapsto$ Ciphertext 'X'.

However, we are told that the code is a Caesar cipher. The numerical equivalent of 'E' is 4 and the numerical equivalent of 'X' is 23. Thus we are led to guess that the Caesar cipher is given by $x \longmapsto x+19 \pmod{26}$. It follows that the cleartext letter = 7+ ciphertext letter (mod 26).

Ciphertext letter	numerical equiv. no.	no. +7 mod 26	Cleartext letter
E	4	11	L
N	13	20	U
V	21	2	C
D	3	10	K
R	17	24	Y

[6]

NB The question does not ask you to decipher the whole ciphertext.

(b) We begin by calculating $\begin{vmatrix} 2 & 3 \\ 7 & 8 \end{vmatrix}$ mod 26.

This is equal to $2 \cdot 8 - 3 \cdot 7 = -5 \equiv 21 \pmod{26}$.

We now use Euclidean algorithm on 21, 26:

$26 = 21 + 5$, $21 = 5 \times 4 + 1$, $5 = 5 \times 1 + 0$.

$\therefore$ gcd(21, 26) = 1 and so 21 is invertible mod 26.

We now compute its inverse.

$1 = 21 - 5 \times 4$; $1 = 21 - (26 - 21) \times 4$,

$1 = 21 - 26 \times 4 + 4 \times 21$; $1 = 5 \times 21 - 26 \times 4$.

It follows that $1 \equiv 5 \times 21 \pmod{26}$ and so $\underline{21^{-1} = 5}$. [2]

We may now compute the deciphering matrix as the inverse mod 26 of our enciphering matrix. This is

$$21^{-1} \begin{pmatrix} 8 & -3 \\ -7 & 2 \end{pmatrix} = 5 \begin{pmatrix} 8 & -3 \\ -7 & 2 \end{pmatrix} = \begin{pmatrix} 40 & -15 \\ -35 & 10 \end{pmatrix} \equiv \begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix}$$

Check: $\begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 105 & 130 \\ 104 & 131 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Deciphering matrix = $\begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix}$ [3]

$$\begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} F \\ W \end{pmatrix} = \begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} 5 \\ 22 \end{pmatrix} = \begin{pmatrix} 0 \\ 19 \end{pmatrix} = \begin{pmatrix} A \\ T \end{pmatrix}$$

$$\begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} M \\ D \end{pmatrix} = \begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} 12 \\ 3 \end{pmatrix} = \begin{pmatrix} 19 \\ 0 \end{pmatrix} = \begin{pmatrix} T \\ A \end{pmatrix}$$

$$\begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} I \\ Q \end{pmatrix} = \begin{pmatrix} 14 & 11 \\ 17 & 10 \end{pmatrix} \begin{pmatrix} 8 \\ 16 \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \end{pmatrix} = \begin{pmatrix} C \\ K \end{pmatrix}$$

The cleartext is therefore $\underline{\text{ATTACK}}$ [2]

(c) We have to calculate 1973^{157} modulo 2773.

We write $157 = 10011101$ in binary.

This enables us to write

$157 = 128 + 16 + 8 + 4 + 1$. [2]

We now construct a table of repeated squares mod 2773.

1973^2	2210
1973^4	847 *
1973^8	1975 *
1973^{16}	1787 *
1973^{32}	1646
1973^{64}	95
1973^{128}	706 *

3

2

[5]

$\therefore 1973^{157} \equiv 1973^{128} \cdot 1973^{16} \cdot 1973^8 \cdot 1973^4 \cdot 1973$

$\equiv 706 \cdot 1787 \cdot 1975 \cdot 847 \cdot 1973$

$\equiv 2680 \ldots 2116 \ldots 894$

$\equiv 234 \pmod{2773}$.

The cleartext is 234.

2(a) In this context, a <u>code</u> is simply a non-empty subset $C \subseteq \mathbb{Z}_2^n$. [1]

If $\underline{x}, \underline{y} \in \mathbb{Z}_2^n$, then the <u>Hamming distance</u> $d(\underline{x}, \underline{y})$ between $\underline{x}$ and $\underline{y}$ is the no. of positions where $\underline{x}$ and $\underline{y}$ differ. [1]

The <u>minimum distance</u> of the code C is the no.

$$d = \min\{ d(\underline{x}, \underline{y}) : \underline{x}, \underline{y} \in C \text{ and } \underline{x} \neq \underline{y} \}$$

(NB) [1]

We prove that if C has min. distance d then it can correct upto t errors where $d \geq 2t+1$.

Let $\underline{x}$ be a codeword and let $\underline{y}$ be the received vector. We suppose that at most t errors have occurred so that

$$d(\underline{x}, \underline{y}) \leq t.$$

Suppose that $\underline{x}'$ is another codeword.

We prove that $d(\underline{y}, \underline{x}') > t$.

Suppose not. Then $d(\underline{y}, \underline{x}') \leq t$.

But then

$$d(\underline{x}, \underline{x}') \leq d(\underline{x}, \underline{y}) + d(\underline{y}, \underline{x}') \leq 2t < 2t+1 = d$$

This contradicts the fact that the min distance is d.

[5]

$C \subseteq \mathbb{Z}_2^n$ is <u>linear</u> if $\underline{0} \in C$ and

$\underline{x}, \underline{y} \in C \Rightarrow \underline{x}+\underline{y} \in C$. [1]

We prove that in a linear code

d = min. distance = min. weight.

Observe that $d(\underline{x}, \underline{y}) = d(\underline{x}-\underline{y}, \underline{0}) = w(\underline{x}-\underline{y})$.

Let $\underline{x}, \underline{y} \in C$ s.t. $d = d(\underline{x}, \underline{y})$.

Then Since C is linear, $\underline{x}-\underline{y}\ (= \underline{x}+\underline{y}) \in C$.

But $w(\underline{x}-\underline{y}) = d$. Thus min. weight

of C is <u>at most</u> d. But if $\underline{z} \in C$

then $w(\underline{z}) = d(\underline{z}, \underline{0})$. Thus min. distance

= min weight. [3]

2 (b)

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

We find a basis for the nullspace of G.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{aligned} u + x + y = 0 &\Rightarrow u = \underset{\alpha}{x} + \underset{\beta}{y} \\ v + x + z = 0 &\Rightarrow v = \underset{\alpha}{x} + \underset{\gamma}{z} \\ w + y + z = 0 &\Rightarrow w = \underset{\beta}{y} + \underset{\gamma}{z} \end{aligned}$$

$$\begin{pmatrix} \alpha+\beta \\ \alpha+\gamma \\ \beta+\gamma \\ \alpha \\ \beta \\ \gamma \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \gamma \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\therefore H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

[6]

The vector 101111 is received. To calculate the most likely codeword transmitted we find

$$H\begin{pmatrix}1\\0\\1\\1\\1\\1\end{pmatrix} = \begin{pmatrix}0\\1\\0\end{pmatrix} \leftarrow 5^{th} \text{ column of } H.$$

This suggests that the 5^{th} bit is in error and the most likely codeword sent was $1011\overset{*}{0}1$

---

3 (a) If $f, g : \mathbb{N} \to \mathbb{N}$ then $f = O(g)$ if $\exists n_0 \in \mathbb{N}$ and a constant C s.t. $\forall n \geq n_0$

$$f(n) \leq C\, g(n). \qquad [1]$$

A problem can be solved in <u>polynomial-time</u> if there is an algorithm that solves it whose time-complexity function is $O(\text{some polynomial})$ [1]

A	B	C	D
n^2	2^n	$\log\log n$	$\log n$

[1]

fastest ⟶ slowest

C $\log\log n$ | D $\log n$ | A n^2 | B 2^n

Let n have d digits in base 10.

Then $10^{d-1} \leq n < 10^d$.

[Thus $d-1 \leq \log_{10}(n) < d$.

Hence $d = \lfloor \log_{10}|n| \rfloor + 1$.] Thus $n \approx 10^d$

No. of trial divisions needed to factorize n is $\sqrt{n}$ (at most).

If $n \approx 10^d$ the time complexity is at worst $d \longmapsto 10^{\frac{d}{2}}$. This is exponential in the length of n. [4]

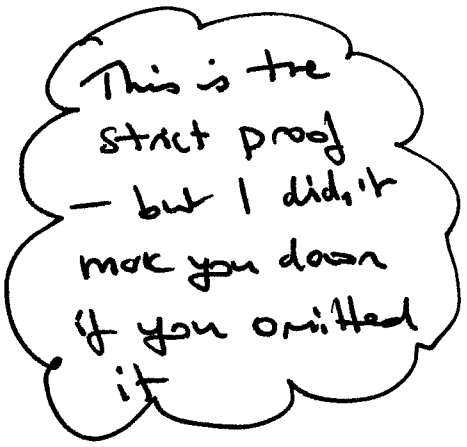
This does <u>not</u> imply that the problem of deciding whether a no. is prime or not is intractable — merely this algorithm. In fact, this problem is known to be <u>intractable</u>. [1]

3(b) Let $n = u^2 - v^2$. Then $n = (u+v)(u-v)$.
Put $a = u+v$ and $b = u-v$. Then $n = ab$, $a \geq b$.

Let $n = ab$ where $a \geq b$.
Put $u = \frac{1}{2}(a+b)$, $v = \frac{1}{2}(a-b)$. Since n is odd, a & b are odd, & so $a+b$, $a-b$ are even.
Thus u and v are whole nos.
Observe that $u^2 - v^2 = ab$. [3]

Given $n = ab$, $a \geq b$ define $u = \frac{1}{2}(a+b)$, $v = \frac{1}{2}(a-b)$
Thus $(a,b) \xmapsto{\alpha} \left(\frac{1}{2}(a+b), \frac{1}{2}(a-b)\right)$

Given $n = u^2 - v^2$ define $a = u+v$, $b = u-v$.
Thus $(u,v) \xmapsto{\beta} (u+v, u-v)$.



- $\beta\alpha(a,b) = \beta\left(\frac{1}{2}(a+b), \frac{1}{2}(a-b)\right)$
  $= (a, b)$
- $\alpha\beta(u,v) = \alpha(u+v, u-v)$
  $= (u, v)$

We factorize 5959. $\sqrt{5959} \approx 77.2$.

Put initial $t = 78$.

t	$t^2 - n$
78	125 ✗
79	282 ✗
80	$441 = 21^2$

$\Rightarrow$ $80^2 - n = 21^2$

$\therefore n = 80^2 - 21^2$

$= \underline{101 \times 59}$ [2]

This technique is most effective when the two factors are not too far apart. [1]

3(c)

PNT: $\lim_{x \to \infty} \dfrac{\pi(x)}{x / \ln(x)} = 1$ [1]

Let p be a prime with 200 decimal digits.

Then $10^{199} < p < 10^{200}$.

$\therefore$ # such primes $\approx \dfrac{10^{200}}{\ln(10^{200})} - \dfrac{10^{199}}{\ln(10^{199})}$

$\approx 1.95 \times 10^{197}$ primes.

We are given that

$$2^n \not\equiv 2 \pmod{n}.$$

FLT says that if p is a prime

then $a^p \equiv a \pmod{p}$ for any a.

We deduce that n is <u>Composite</u>. [3]

4. We proof that U_p, when p is a prime, is cyclic. This is a complex proof so the marking was generous and I was mainly looking for good ideas. To ease notation, put $G = U_p$ and $|G| = n$.

By Lagrange's theorem, the order of each element of G divides n.

For each $d \mid n$, define $X_d \subseteq G$ to consist of all elements whose order is exactly d.

We have that

- $G = \bigcup_{d \mid n} X_d$
- $X_d \cap X_{d'} = \emptyset$ if $d \neq d'$
- Some X_d could be empty.

It follows that

$$n = \sum_{d \mid n} |X_d|$$

We shall prove the following:

|| if there is an element of order d then there are exactly $\phi(d)$ elements of order d. || (*)

Thus if $X_d \neq \emptyset$ then $|X_d| = \phi(d)$.

BUT we are given that $\sum_{d|n} \phi(d) = n$.

Thus none of the sets X_d can be empty.

In particular, there must be $\phi(n)$ elements of order n.

It follows that G is cyclic.

[5]

We now prove the claim (*)
i.e.

Let a be an element of order d dividing n.
We shall prove that there are exactly $\phi(d)$ elements of order d. Specifically, we shall prove that the elements

$$a^j \quad \text{where} \quad 1 \leq j \leq d-1, \quad \gcd(j,d) = 1$$

are precisely all the elements of order d.

$$\langle a \rangle = \{ a, a^2, \ldots, a^{d-1}, a^d = 1 \}.$$

By Lagrange's theorem applied to $\langle a \rangle$ all elements of $\langle a \rangle$ have orders dividing d. The polynomial $x^d - 1$ over $\mathbb{Z}_p$ has at most d roots. Thus the elements of $\langle a \rangle$ are all the roots of $x^d - 1$. But any element of order d is a root of $x^d - 1$. It follows that all elements of order d must live in $\langle a \rangle$. [5]

- Let $\gcd(d, i) = d' > 1$.

Then the order of a^i is strictly less than d.

Put $d = d'j$, $i = d'k$ some j, k.

Now $(a^i)^j = a^{ij}$ and $ij = d'kj = d'jk = dk$

$\therefore d \mid ij$. Thus $(a^i)^j = 1$. It follows that the order of a^i is $\leq j < d$. [2]

- Let $\gcd(d, i) = 1$. We prove that the order of a^i is d. Suppose $(a^i)^{d'} = 1$ where $d' < d$.

Then $a^{id'} = 1$. Thus $d \mid id'$. $\therefore d \mid d'$ #. [2]

From the proof, we see that U_p has $\varphi(p-1)$ generators. [1] 17

We now find all generators of U_{11}.

We expect there to be $\varphi(10) = 4$ of them.

We try 2 first.

2^1	2^2	2^3	2^4	2^5	2^6	2^7	2^8	2^9	2^{10}
2	4	8	5	10	9	7	3	6	1

Thus 2, 8, 7, 6 are the generators of U_{11}. [5]

5(a)

A <u>free monoid</u> on the set A is the set of all finite strings over A equipped with the binary operation of concatenation. The identity is the empty string ε. [1]

$C \subseteq A^*$ a finite subset is called a <u>code</u> if it generates a free subsemigroup, meaning that each element of C^+ can be written in exactly one way as a product of elements in C. [1]

$C \subseteq A^*$ is called a <u>prefix code</u> if

$x, y \in C$ and $x = yz$ for some $z \Rightarrow z = \varepsilon$.

(No element is a proper prefix of any other). [1]

Suppose $x_1 \ldots x_m = y_1 \ldots y_n$ where $x_i, y_i \in C$ a prefix code.

Compare x_1 and y_1. If $|x_1| = |y_1|$ then $x_1 = y_1$ & cancel.
If $|x_1| > |y_1|$ then y_1 is a proper prefix of x_1 — Contradiction.
Ditto $|y_1| > |x_1|$. Deduce in this way $m = n$ and $x_i = y_i$, as required. [4]

5(b) Given frequencies to probabilities

letter	probability
a	5/48
b	6/48
c	6/48
d	11/48
e	20/48

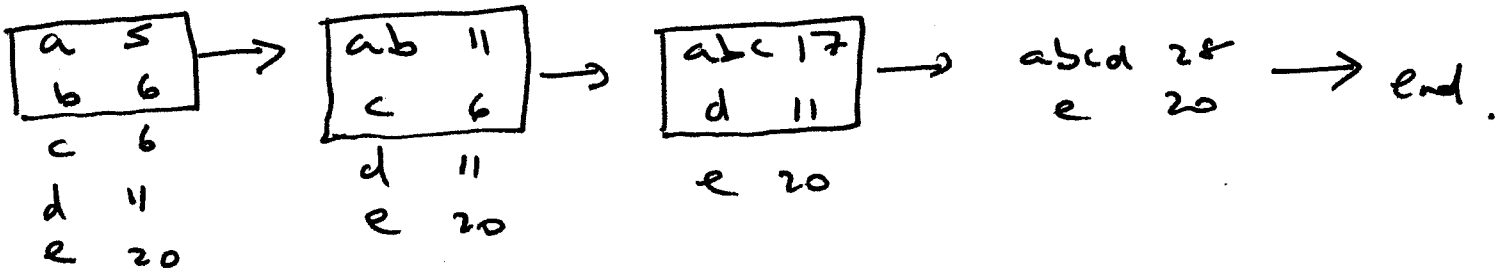
Let $f = (p_1, \ldots, p_m)$.

The binary entropy is given by

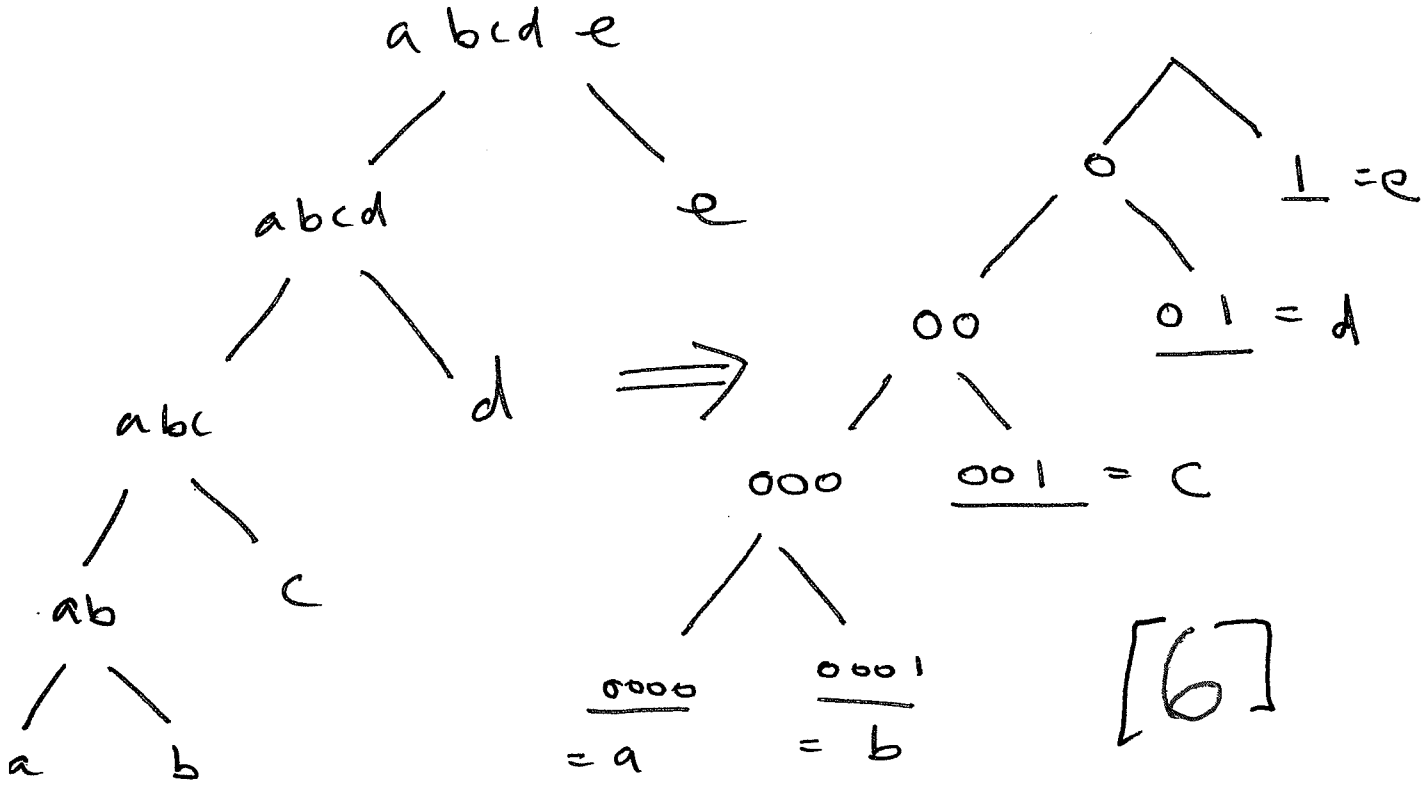
$$H_2(f) = \frac{1}{\log_{10}(2)} \sum_{i=1}^{m} p_i \log_{10}\left(\frac{1}{p_i}\right)$$

Here this equates to $\approx 2 \cdot 10326\ldots$ [4]

We now calculate the Huffman encoding.



This calculation is now re-expanded as a binary tree



[6]

letter	Huffman coding
a	0000
b	0001
c	001
d	01
e	1

There are a number of other possible solutions.

We now calculate the average word length. [2]

$$L = 0.1042 \times 4 + 0.125 \times 4 + 0.125 \times 3 + 0.2292 \times 2 + 0.4167 \times 1$$

$$= \underline{2.167}$$

We have $L \geq H$ (just!) [1]