EEE605 – DATA COMPRESSION HOMEWORK – 2

1) a) Given the probability model in Table, find the real valued tag for the sequence *a1 a1 a3 a2 a3 a1*.

b) Again using the table, decode a sequence of length 10 with the tag 0.63215699.

Probability Model	
Letter	**Probability**
a_1	.2
a_2	.3
a_3	.5

2) Given the frequency counts shown in Table:

a) What is the word length required for unambiguous encoding?

b) Find the binary code for the sequence *abacabb*.

c) Decode the code you obtained to verify that your encoding was correct.

Frequency Counts	
Letter	**Count**
a	37
b	38
c	25

3) Given an initial dictionary consisting of the letters a b r y ƀ, encode the following message using the LZW algorithm: aƀbarƀarrayƀbyƀbarrayarƀbay.

4) A sequence is encoded using the LZW algorithm and the initial dictionary shown in Table. The output of the LZW encoder is the following sequence: *3 1 4 6 8 4 2 1 2 5 10 6 11 13 6*

Initial Dictionary	
Index	**Entry**
1	*a*
2	·
3	*r*
4	*t*

Decode this sequence.

5) Encode the following sequence using the LZ77 algorithm: *barrayar ƀbar ƀby ƀbarrayar ƀbay*

Assume you have a window size of 30 with a look-ahead buffer of size 15.

Furthermore, assume that *C(a) = 1*, *C(b) = 2*, *C(ƀ) = 3*, *C(r) = 4*, and *C(y) = 5*.

6) A sequence is encoded using the LZ77 algorithm. Given that *C(a)* = 1, *C(*ƀ*)* = 2, *C(r)* = 3, and *C(t)* = 4,

decode the following sequence of triples: *(0,0,3) (0,0,1) (0,0,4) (2,8,2) (3,1,2) (0,0,3) (6,4,4) (9,5,4)*. Assume that the size of the window is 20 and the size of the look-ahead buffer is 10. Encode the decoded sequence and make sure you get the same sequence of triples.

7) Given the following primed dictionary and the received sequence below, build an LZW dictionary *and* decode the transmitted sequence.

Received Sequence: *4, 5, 3, 1, 2, 8, 2, 7, 9, 7, 4*

Initial dictionary:

Index	Entry
1	S
2	$\not b$
3	I
4	T
5	H

8) Given the sequence *eta ƀceta ƀand ƀbeta ƀceta:*

(a) Encode using the Burrows-Wheeler transform and move-to-front coding.

(b) Decode the encoded sequence.

9)

A sequence is encoded using the LZW algorithm and the initial dictionary shown in the following table.

Index	Entry
1	a
2	c
3	r
4	y
5	Δ

The output of the LZW encoder is the following sequence: 1, 5, 2, 1, 3, 5, 9, 3, 1, 4, 7. Please decode the message.

10)

We try to encode the sequence $cat\Delta ate\Delta hat$ using *ppma* with maximal context length $N = 1$ and an integer arithmetic code with a word length of 6. The alphabet set is $\{h, e, t, a, c, \Delta\}$. Assume that we have finished encoding of $cat\Delta ate\Delta$ and obtained the following context table:

order	context	symbol occurrence counts							Total
		h	e	t	a	c	Δ	<ESC>	
1	c	0	0	0	1	0	0	1	2
	a	0	0	2	0	0	0	1	3
	t	0	1	0	0	0	1	1	3
	Δ	0	0	0	1	0	0	1	2
	e	0	0	0	0	0	1	1	2
0	/	0	1	2	2	1	2	1	9
−1	/	1	1	1	1	1	1	0	6

Note that in the context table, the cumulative count is calculated from left-to-right. For example, the zero-order context has cumulative count as follows: $h = 0$, $e = 1$, $t = 3$, $a = 5$, $c = 6$, and $\Delta = 8$.

The transmitted sequence after encoding of $cat\Delta ate\Delta$ is 11000011111101100000100 and the current lower and upper bounds are $l = 011100$ and $u = 110011$. Please encode the next letter h and write down the newly transmitted bits for h and the updated lower and upper bounds.

Hint: For an integer AC implementation, the message interval can be updated by:

$$l^{(n)} = l^{(n-1)} + \lfloor (u^{(n-1)} - l^{(n-1)} + 1) \times cum\_count(x_n - 1)/total\_count \rfloor,$$

$$u^{(n)} = l^{(n-1)} + \lfloor (u^{(n-1)} - l^{(n-1)} + 1) \times cum\_count(x_n)/total\_count \rfloor - 1.$$