Mitderm I
EE499/599: Data Compression and Information Theory
5/11/06
Prof. Thinh Nguyen

You have 80 minutes to complete the quiz with 28 points total. When in doubt, clearly state your assumptions. You will be given partial credits for showing your work. Good luck!

Problem 1: (8 pts)

A binary communication system makes use of the symbols "zero" and "one". As a result of distortion, errors are sometimes made during transmission. Consider the following events:

U0: a "zero" is transmitted;
U1: a "one" is transmitted;
V0: a "zero" is received;
V1: a "one" is received.

The following probabilities are given:
P(U0) = ½, P(V0/U0) = ¾, P(V0/U1) = ½.

a) How much information do you receive when you learn which symbol has been received, while you know that a "zero" has been transmitted? (2pts)

We have P(V1|U0) = 1 – P(V0|U0) = ¼
Thus for the uncertainty with regard to the received symbol, given that a "zero" has been transmitted, we find

H(V|U0) = -P(V0|U0)log (P(V0|U0)) – P(V1|U0)log P(V1|U0) = -3/4 log(3/4) – ¼ log (1/4) = 0.82 bit.

b) How much information do you receive when you learn which symbol has been received, while you know which symbol has been transmitted? (2pts)

We have P(U0, V1) = 1/8, P(U1, V0) = ¼ and P(U1, V1) = ¼

For the amount of information with regard to the received symbol, given the transmitted symbol, it now follows that

$$H(V|U) = -\sum_{i=0}^{1}\sum_{j=0}^{1} p(u_{i,} v_i)\, p(v_j \mid u_j) = -3/8(\log 3/4) - 1/8\log(1/4) - 1/4\log(1/2) - 1/4\log(1/2) = 0.91 bit$$

c) Determine the amount of information that you receive when someone tells you which symbol has been transmitted and which symbol has been received. (2pts)

$$H(V|U) = -\sum_{i=0}^{1}\sum_{j=0}^{1} p(u_{i,}v_i)\, p(v_j, u_j) = 1.91 bits$$

d) Determine the amount of information that you receive when someone tells you which symbol has been transmitted, while you know which symbol has been received. (2pts)

P(V0) = 5/8 and P(V1) = 3/8
H(V) = -5/8 log (5/8) – 3/8 log (3/8) = 0.96 bit
H(U|V) = H(U,V) - H(V) = 1.91 – 0.96 = 0.95 bit.

Problem 2: Suppose you have a fax image consists of lines, each line has 50 pixels, roughly 10 black pixels and 40 white pixels. (14pts)

a) Provide a “good” divisor m to be used for Golomb coding this fax image. (1pts)

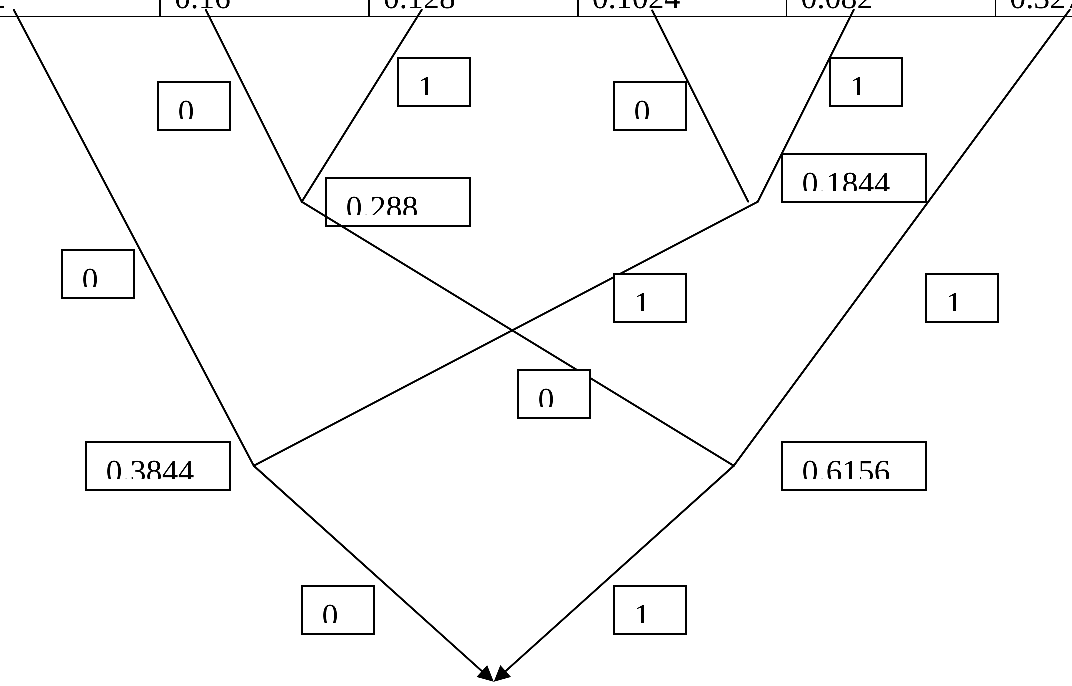
$$m = \left\lceil -\frac{1}{\log_2 P(0)} \right\rceil = \left\lceil -\frac{1}{\log_2 0.8} \right\rceil = 4$$

b) Provide the Golomb code for the five symbols: 1, 01, 001, 0001, and 00001 using the value of m above (3pts).

Input	1	01	001	0001	00001
Output	000	001	010	011	1000

c) What is the Huffman code for these five symbols? (3pts)

1	01	001	0001	00001	00000xxxx
0.2	0.16	0.128	0.1024	0.082	0.3276



1	01	001	0001	00001	00000xxx
00	100	101	010	011	11

d) Suppose a line in the fax image contains 50 symbols, and you want to use the integer arithmetic coding to code the entire line, what is the minimum number of bits required to code the line so that every line is distinguishable? Show steps by steps of the encoding process for the first 4 pixels in this order: black, white, white, black. You don't need to send the termination tag (5pts).

m = ceiling((log 200)) = 8.

L = 0, R = 255
Total_Count = 50
Cum_Count(0) = 40
Cum_Count(1) = 50

Encode black pixel:
L = 0 + floor[(256)(0)/50] = 0 = 00000000, R = 0 + floor[(256)(1)/50] – 1 = 50 = 00110010
Lower half -> output 0
L = 00000000, R = 01100101

Output 0:
L = 00000000,   R = 11001011 = 203

Encode white pixel:

L = floor[(204)(10)/50] = 40 = 00101000, R = (204)(50)/50 -1 = 203 = 11001011.

Encode white pixel:
W = 164
L = 40 + floor[(164x10/50)] = 72 = 01001000, R = 40 + (164)(50)/50 -1 203 11001011

Encode black pixel:
L = 72 = 01001000, R = 72 + floor((132)(10)/50) -1 = 97 = 0110 0001
Output 0:
L = 10010000, R = 110000011

Output 1:
L = 00100000, R = 10000111

Output Sequence: 0001

e) Suppose you come up with a new compression algorithm to code a sequence consisting 0 and 1 by coding the runlength, i.e., 1, 01, 001, … You also prove that the average code length of these runlengths using your new algorithm is $2p^2$ where p is the probability of 1 happens.   What is the range of p in order for your compression algorithm to be useful? (2pts).

$$2p^2/(1/p) < 1 \rightarrow p < 1/\sqrt[3]{2}$$

Problem 3: (6pts)
This problem deals with the dictionary coding techniques:

A B A C D A E F	A C A E A B A

Search Buffer   Look-ahead Buffer

a) Assuming the search buffer size is 8 and the initial contents are shown in the figure above, show the transmitted code for the next 7 symbols in the look-ahead buffer using LZ77 (3 pts).

ABACDAEF ACAEABA
Send: 62A

CDAEFACA EABA
Send: 51A

AEFACAEA BA
Send: 00B

AEFACAEA A
11 “end symbol”

b) Assume the primary symbols are A, B, C, D, E, F. Show the dictionaries at the encoder and decoder using LZW after the encoding and decoding the following sequence of symbols: A, B, A, C, D, F, A, B.

Encoder:	Decoder:
A	A
B	B
C	C
D	D
E	E
F	F
AB	AB
BA	BA
BC	BC
CD	CD
DF	DF
FA	