CSC 310S
Spring, 2003

January 13, 2003
DUE: January 29, 2003

Information Theory

Assignment #1 --- Sample Answers

1. [10 marks]
Characterise each of the five codes below as being one of *Instantaneous*, *Uniquely Decipherable* (but not Instantaneous), or *Ambiguous*. In each case, provide sufficient evidence to justify your characterization.

msg	code (i)	code(ii)	code (iii)	code (iv)	code (v)
a	11	0100	100	011	00
b	101	1011	001	11	011
c	001	0101	1010	1011	110
d	100	11	1000	110	10
e	0101	110	11	0100	010
f	0001	0111	101	0110	111

(i).

$C_1 = \phi$
$C_\infty = \phi$ So, INSTANTANEOUS

(ii).

$C_1 = 0$
$C_2 = 100, 101, 111$
$C_3 = 1$
$C_4 = 011, 1, 10$
$C_5 = 011, 1, 10, 11$
$\{11\} \ \varepsilon \ C_0 \cap C_\infty$ So, AMBIGUOUS

(iii).

$C_1 = 0$
$C_2 = 01$
$C_3 = \phi$
$C_0 \cap C_\infty = \phi$ So, UNIQUELY DECODABLE

(iv).

$C_1 = 0$
$C_2 = 11, 100, 110$
$\{ 11, 110 \} \ \varepsilon \ C_0 \cap C_\infty$ So, AMBIGUOUS

(v).

$C_1 = \phi$
$C_\infty = \phi$ So, INSTANTANEOUS

2. [20 marks]
Consider the following game: I take a penny, a nickel, and a dime from my stash of coins, and flip each of them. You win each coin that comes up "tails". (Assume that each of them is a "fair" coin, i.e., equally likely to come out either "heads" or "tails" when flipped.)

Define the random variables as follows:

P (= h or t) is the outcome of the penny
N (= h or t) is the outcome of the nickel
D (= h or t) is the outcome of the dime
T (= 0, 1, 2, or 3) is the total numbers of "tails"
W (= 0, 1, 5, 6, 10, 11, 15, or 16) is your total winnings

Calculate each of the information quantities below, and justify your calculations where necessary. (Note that only a few of these quantities need to be calculated from their base definitions. Many of them can instead be derived from prior answers by applying some of the Information Theory "Facts" presented in class.)

a. $H(P) = \log_2(2) = 1$ bit
b. $H(T) = H(\frac{1}{8},\frac{3}{8},\frac{3}{8},\frac{1}{8}) = 1.8113$ bits
c. $H(W) = \log_2(8) = 3$ bits
d. $H(N,D) = H(N)+H(D) = 2$ bits
e. $H(N,T) = H(\frac{1}{8},\frac{2}{8},\frac{1}{8},\frac{1}{8},\frac{2}{8},\frac{1}{8}) = 2.5$ bits
f. $H(P,W) = H(W) = 3$ bits
g. $H(T,W) = H(W) = 3$ bits
h. $H(N \mid P) = H(N) = 1$ bit
i. $H(N \mid T) = \frac{3}{4} \times H(\frac{1}{3},\frac{2}{3}) = .6887$ bits
j. $H(W \mid P) = H(N,D) = 2$ bits
k. $H(T \mid W) = H(T,W)-H(W) = 0$ bits
l. $H(T \mid N) = H(\frac{1}{4},\frac{2}{4},\frac{1}{4}) = 1.5$ bits
m. $H(P \mid W) = H(P,W)-H(W) = 0$ bits
n. $H(W \mid T) = \frac{3}{4} \times H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) = 1.1887$ bits
o. $I(N,P) = H(N)-H(N \mid P) = 0$ bits
p. $I(N,T) = H(T)-H(T \mid N) = .3113$ bits
q. $I(W,P) = H(W)-H(W \mid P) = 1$ bit
r. $I(T,W) = H(W)-H(W \mid T) = 1.8113$ bits
s. $H(W \mid \{P,N\}) = H(D) = 1$ bit
t. $H(W \mid \{D,T\}) = \frac{1}{2} \times H(\frac{1}{2},\frac{1}{2}) = .5$ bits

3. [10 marks]
For each of the following combinations of a code alphabet base and a set of code word lengths, either (1) produce an *instantaneous* code in the base with the specified code word lengths, or else (2) demonstrate why this is not possible.

case	code alphabet	code word lengths
(i)	base 2 {0, 1}	1, 2, 3, 4, 5, 5, 5, 5
(ii)	base 2 {0, 1}	2, 2, 3, 3, 3, 4, 5, 5
(iii)	base 3 (0, 1, 2}	1, 1, 2, 3, 3, 3, 3, 3
(iv)	base 4 (0, 1, 2, 3}	1, 1, 1, 2, 2, 2, 2, 2
(v)	base 5 (0, 1, 2, 3, 4}	1, 1, 1, 1, 2, 2, 2, 2

(i).

$$\frac{1}{32}(16+8+4+2+4\times1) = \frac{34}{32} > 1$$

Therefore, no instantaneous code exists.

(ii).

$$\frac{1}{32}(2\times8+3\times4+1\times2+2\times(1)) = \frac{32}{32} \leq 1$$

Therefore, an instantaneous code exists.
Example: { 00, 01, 100, 101, 110, 1110, 11110, 11111 }

(iii).

$$\frac{1}{27}(2\times9+1\times3+5\times1) = \frac{26}{27} = \frac{26}{27} \leq 1$$

Therefore, an instantaneous code exists.
Example: { 0, 1, 20, 210, 211, 212, 220, 221 }

(iv).

$$\frac{1}{16}(3\times4+5\times1) = \frac{17}{16} > 1$$

Therefore, no instantaneous code exists.

(v).

$$\frac{1}{25}(4\times5+4\times1) = \frac{24}{25} = \frac{24}{25} \leq 1$$

Therefore, an instantaneous code exists.
Example: { 0, 1, 2, 3, 40, 41, 42, 43 }

4. [20 marks]
Recent statistics indicate that the populations of Canada's provinces and territories are as follows:

Province	*Population*	*Province*	*Population*
Alberta	3,064,249	Ontario	11,874,436
British Columbia	4,095,934	Prince Edward Island	138,514
Manitoba	1,150,034	Quebec	7,410,504
New Brunswick	757,007	Saskatchewan	1,015,783
Newfoundland	533,761	Northwest Territories	40,860
Nova Scotia	942,391	Yukon	29,885
Nunavut	28,159		

A(i). How much information is conveyed by learning the province or territory of residence of a randomly selected Canadian resident?

H(Which Province?) = H(.0986,.1318,.0370,.0244,.0172,.0303,.0009,.3819,.0045,.2384,.0327,.0013,.0010)

H(Which Province?) = 2.5259 bits

A(ii). How much information is conveyed by learning whether or not a randomly selected Canadian resident lives in Ontario?

H(Ontario resident?) = H(.3819,.6181) = .9594 bits

A(iii). How much information is conveyed by learning whether or not a randomly selected Canadian resident lives in the Yukon?

H(Yukon resident?) = H(.00096,.99904) = .0110 bits

B. Determine *binary*, *ternary*, and *decimal* Huffman Codes for encoding the Province or territory of residence of each member of a large group of Canadian residents, and calculate and compare their *relative redundancies*.

province	population	prob	base 2	base 3	base 10	base 16
ONT	11,874,436	.3819	0	0	0	0
QUE	7,410,504	.2384	10	1	1	1
BC	4,095,934	.1318	1110	20	2	2
ALB	3,064,249	.0986	1101	21	3	3
MAN	1,150,034	.0370	11100	220	4	4
SAS	1,015,783	.0327	11110	221	5	5
NS	942,391	.0303	11111	2220	6	6
NB	757,007	.0244	111010	2221	7	7
NF	533,761	.0172	1110110	22220	8	8
PEI	138,514	.0045	11101110	22221	90	9
NWT	40,860	.0013	111011110	222220	91	A
YUK	29,885	.0010	1110111110	222221	92	B
NUN	28,159	.0009	1110111111	222222	93	C

$L_2(PROV)$ = 2.614	$L_3(PROV)$ = 1.637	$L_{10}(PROV)$ = 1.008	$L_{16}(PROV)$ = 1.000
$H_2(PROV)$ = 2.527	$H_3(PROV)$ = 1.594	$H_{10}(PROV)$ = .761	$H_{16}(PROV)$ .641
Rel. Red.: 3.3%	Rel. Red.: 2.6%	Rel. Red.: 24.5%	Rel. Red.: 37.8%