CSC 2508S
Spring, 2002

September 18, 2002
DUE: October 2, 2002

Information: Quantification, Representation and Manipulation

Assignment #1

1. [30 marks]

Consider a game in which one player flips a fair coin while the other rolls a fair die. Denote the outcome of the coin by $C \varepsilon \{H,T\}$, and the outcome of the die by $D \varepsilon \{1,2,3,4,5,6\}$. Let the "scores" of the two players be denoted by S_1 and S_2, respectively.

Define the scores obtained by the two players by:

$$S_1 = \begin{cases} 6 & \text{if } C=H \\ 1 & \text{if } C=T \end{cases} \qquad \text{and} \qquad S_2 = D$$

and the "winning" variable, W, to be:

$$W = \begin{cases} +1 & \text{if } S_1 > S_2 \\ 0 & \text{if } S_1 = S_2 \\ -1 & \text{if } S_1 < S_2 \end{cases}$$

Evaluate each of the following quantities using the "facts" presented in lecture (which make many of the parts trivial to calculate). Indicate the basis for each answer.

i. $H(D) = \log_2(6) = 2.5850$ bits

ii. $H(C) = \log_2(2) = 1$ bit

iii. $H(W) = H(\frac{5}{12},\frac{1}{6},\frac{5}{12}) = 1.4833$ bits

iv. $H(W \mid D) = \log_2(2) = 1$ bit

v. $H(W \mid C) = H(\frac{1}{6},\frac{5}{6}) = .6500$ bits

vi. $H(D \mid W) = \frac{5}{6}\log_2(5)+\frac{1}{6}\log_2(2) = 2.1016$ bits

vii. $H(C \mid W) = \frac{5}{6}\log_2(1)+\frac{1}{6}\log_2(2) = .1667$ bits

viii. $H(W,D) = \log_2(12) = 3.5850$ bits

ix. $H(W,C) = H(\frac{5}{12},\frac{1}{12},\frac{1}{12},\frac{5}{12}) = 1.6500$ bits

x. $I(D,C) = H(D)-H(D \mid C) = 0$ bits

xi. $I(D,W) = H(D)-H(D \mid W) = .4834$ bits

xii. $I(C,W) = H(C)-H(C \mid W) = .8333$ bits

xiii. $H(W \mid (C,D)) = 0$ bits

xiv. $H(C \mid (W,D)) = 0$ bits

xv. $H(D \mid (W,C)) = \frac{5}{6}\log_2(5) = 1.9349$ bits

b. For the fifteen parts of a, how do the answers change if each "C" is replaced by "S_1", and each "D" is replaced by "S_2"? Explain your answer.

The answers do not change at all since $I(C,S_1)=H(C)=H(S_1)$, and also $I(D,S_2)=H(D)=H(S_2)$. That is, the pairs are each fully equivalent from an information bearing point of view.

c. How do the answers to part a change if the definition of S_1 is changed to be:

$$S_1 = \begin{cases} 5 & \text{if } C=H \\ 2 & \text{if } C=T \end{cases}$$

Parts i, ii, iii, x, and xiii do not change. The others change as follows:

iv. $H(W \mid D) = \frac{2}{3}\log_2(2) = .6667$ bits

v. $H(W \mid C) = H(\frac{2}{3},\frac{1}{6},\frac{1}{6}) = 1.2517$ bits

vi. $H(D \mid W) = \frac{5}{6}H(\frac{2}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5})+\frac{1}{6}H(\frac{1}{2},\frac{1}{2}) = 1.7683$ bits

vii. $H(C \mid W) = \frac{5}{6}H(\frac{4}{5},\frac{1}{5})+\frac{1}{6}H(\frac{1}{2},\frac{1}{2}) = .7683$ bits

viii. $H(W,D) = H(\frac{1}{6},\frac{1}{6},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12}) = 3.2516$ bits

ix. $H(W,C) = H(\frac{1}{3},\frac{1}{3},\frac{1}{12},\frac{1}{12},\frac{1}{12},\frac{1}{12}) = 2.2516$ bits

xi. $I(D,W) = H(D)-H(D \mid W) = .8166$ bits

xii. $I(C,W) = H(C)-H(C \mid W) = .2317$ bits

xiv. $H(C \mid (W,D)) = \frac{1}{3}\log_2(2) = .3333$ bits

xv. $H(D \mid (W,C)) = \frac{2}{3}\log_2(4) = 1.3333$ bits

2. [30 marks]
The populations of thirteen cities in Canada's provinces and territories are as follows:

City	*Population*	*City*	*Population*
Edmonton, ALB	917,500	Yellowknife, NWT	37,360
Vancouver, BC	1,995,900	Toronto, ONT	4,594,900
Winnepeg, MAN	676,400	Charlottetown, PEI	32,531
St. John, NB	127,300	Montreal, PQ	3,428,300
St. John's, NF	173,600	Saskatoon, SASK	229,300
Halifax, NS	348,000	Whitehorse, YUK	31,070
Pond Inlet, NUN	1,154		

a) How much information is conveyed by learning the city of residence of a randomly selected person who lives in one of the 13 cities?

$$H(CITY) = H(.0729,.1585,.0537,.0101,.0138,.0276,.0001,.0030,.3648,.0026,.2722,.0182,.0025) = 2.4356 \text{ bits}$$

b) How much information is conveyed by learning whether or not a randomly selected person lives in Montreal?

$$H(CITY{=}Montreal) = H(.2722,.7278) = .8446 \text{ bits}$$

c) How much information is conveyed by learning whether or not a randomly selected person lives in the Yellowknife?

$$H(CITY{=}Yellowknife) = H(.0030,.9970) = .0295 \text{ bits}$$

d) Determine a *binary* Huffman Code for encoding the city of residence of each member of a large group of people who live in one of the 13 cities.

e) Determine a *ternary* Huffman code for the same situation.

f) Determine a *decimal* Huffman code for the same situation.

g) Calculate and compare the relative redundancy of the three codes in parts (d), (e), and (f).

city	population	prob	base 2	base 3	base 4	base 10	base 16
Toronto	4,594,900	.3648	0	0	0	0	0
Montreal	3,428,300	.2722	10	1	1	1	1
Vancouver	1,995,900	.1585	110	20	2	2	2
Edmonton	917,500	.0729	11100	21	30	3	3
Winnepeg	676,400	.0537	11101	220	31	4	4
Halifax	348,000	.0276	111110	221	32	5	5
Saskatoon	229,300	.0182	111100	2220	330	6	6
St. John's	173,600	.0138	111101	2221	331	7	7
St. John	127,300	.0101	1111110	22220	332	8	8
Yellowknife	37,360	.0030	11111110	22221	3330	90	9
Charlottetown	32,531	.0026	111111110	222220	3331	91	A
Whitehorse	31,070	.0025	1111111110	222221	3332	92	B
Pond Inlet	1,154	.0001	1111111111	222222	3333	93	C
$L_D(CITY) =$			2.5194	1.5684	1.2630	1.0082	1.000
$H_2(CITY) =$			2.4356	1.5367	1.2178	.7332	.6089
Rel.Redund. =			3.33%	2.02%	3.58%	27.3%	39.0%