February 28th, 2003 **Spring Term, 2003**

CSC 310S: Information Theory
Mid-Term Test

NAME _____**KCS**_____ **STUDENT #** _____

#1	#2	#3	#4	#5	TOTAL
/20	/20	/20	/20	/20	/100

[HINT: Try to answer all questions before spending too much time on any one.
Open Book; open notes; calculators allowed.]

1.

A. [8 marks]

A simple model of a source with alphabet $\{S_1, S_2, \ldots, S_q\}$, with respective probabilities $\{p_1, p_2, \ldots, p_q\}$ includes assumptions that:

i. each symbol is generated *independently*,
ii. the probabilities are *static*, and
iii. the probabilities are *known*.

Some of the variants of *Huffman coding* shown in the table below are intended work effectively despite a violation of one of the assumptions above. For each variant, put a check mark in the box corresponding to the assumption (if any) to which the variant is insensitive.

	(i) dependence	(ii) dynamic probs.	(iii) unknown static probs.
Huffman coding of individual symbols			
Huffman coding of groups of symbols	XXX		
Adaptive Huffman coding without "forgetting"			XXX
Adaptive Huffman coding with "forgetting"		XXX	
Huffman coding of Markov model transitions	XXX		

B. [12 marks]

[NOTE: The answers to the four parts of this question should be given at a sufficiently high-level (i.e., no details) that they fit easily in the space provided.]

Assume that, for a symbol alphabet of q symbols, you are given codewords in base r: $\{w_1, w_2, \ldots, w_q\}$, with respective lengths: $\{l_1, l_2, \ldots, l_q\}$. Explain briefly how you would ...

(i). determine whether or not the given code is *instantaneous*.

If the code has the *prefix property* (i.e., no code word is the prefix of another code word), then the code is instantaneous; Otherwise, it is not.

(ii). determine whether or not the given code is *uniquely decodable*.

If the set of code words does not intersect with the set C_∞, as calculated in accordance with the Sardinas-Patterson theorem (see section 1.2 of Jones and Jones), then the code is uniquely decodable; Otherwise, it is ambiguous.

If the given specific code is found to be *ambiguous*, explain briefly how you would ...

(iii). determine whether or not there exists *any instantaneous* code in base r with the specified respective code word lengths.

By Kraft's theorem, if the code word lengths, $\{l_1, l_2, \ldots, l_q\}$, satisfy the inequality:

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq 1$$

then there exists an instantaneous code with those code word lengths; Otherwise, there does not.

(iv). determine whether or not there exists *any uniquely decodable* code in base r with the specified respective code word lengths.

By McMillan's theorem, if the code word lengths, $\{l_1, l_2, \ldots, l_q\}$, satisfy the inequality:

$$\sum_{i=1}^{q} \frac{1}{r^{l_i}} \leq 1$$

then there exists any an uniquely decodable code with those code word lengths; Otherwise, there does not.
(i.e., exactly the same condition as in part (iii).)

2.

A. [8 marks] Consider a source alphabet with *entropy* $H(X)$, and an arbitrary code for that alphabet, which has *average codeword length, L(X)*. Beside each *type* of code in the left column below, enter the *letter* from the column on the right associated with the *tightest* range in which $L(X)$ is guaranteed to lie.

Code Type	Answer	Range	
(i). ambiguous	___a___	a.	$0 \le L(x) < \infty$
		b.	$0 \le L(x) < H(X)+1$
(ii). uniquely decodable	___d___	c.	$0 \le L(x) < H(X)$
		d.	$H(X) \le L(x) < \infty$
(iii). instantaneous	___d___	e.	$H(X) \le L(x) < H(X)+1$
		f.	$H(X)+1 \le L(x) < \infty$
(iv). Huffman code	___e___		

B. [6 marks] Assume that a source generates the digits {1, 2, 3, 4, 5, 6, 7, 8, 9} with independent probabilities respectively (.20, .20, .20, .20, .04, .04, .04, .04, .04}, and you wish to find a *zero redundancy* code for this source.

(i). What base, r, should be chosen for the code?

The base must be $r = 5$.

(ii). Why is the base chosen in part (i) the appropriate choice?

Because the probabilities $.2 = \frac{1}{5}$ and $.04 = \frac{1}{25}$ are each inverse powers of 5.

(iii). Give the nine codewords of a zero-redundancy code in the given base:

1 - 0	4 - 3	7 - 42
2 - 1	5 - 40	8 - 43
3 - 2	6 - 41	9 - 44

C. [6 marks] For a source with alphabet {Y , N} and entropy $H(p_Y, p_N)$, describe briefly a way of constructing a code such that its codeword length is guaranteed to satisfy $L(X) < H(X) + \frac{1}{2}$.

Use grouping into pairs, so that the code words are {YY, YN, NY, NN} with respective probabilities $\{p_Y{}^2, p_Y p_N, p_Y p_N, p_N{}^2\}$. Then, either produce a Huffman code, or simply produce an instantaneous code with code word lengths chosen to be *ceiling* $(-\log_r(p_i p_j))$ for the pair, ij.

3.

A. [10 marks]
For a source with alphabet {a, b, c, d} and respective probabilities {.4, .3, .2, .1},

(i). Give an *instantaneous* code in base 2 with codeword lengths chosen by the Shannon criterion ($l_i = -\log_2(p_i)$):

a - 00	b - 01	c - 100	d - 1100

(ii). Give a base 2 *Huffman code* for the same source:

a - 0	b - 10	c - 110	d - 111

(iii). How many *distinct* Huffman codes exist for this source alphabet with these probabilities?

Number of distinct Huffman codes: $2^3 = 8$
(since the Huffman tree has 3 internal nodes)

B. [10 marks]
Consider using *Arithmetic Coding* to encode in bits (base 2) a source with alphabet {H , T} and respective probabilities {.75, .25}. (Let H correspond to the interval [0.0, .75), and T correspond to the interval [.75, 1.0).)

(i). If the source sequence starts off with "HHH ... ", how many initial bits of the encoded string are known, and what are they?

Number known: 1	Known bit(s): "0"

(ii). If the source sequence starts off with "TTT ... ", how many initial bits of the encoded string are known, and what are they?

Number known: 6	Known bit(s): "111111"

(iii). If the encoded value of a source string is .5 (or .1000000... as a binary fraction), what are the first three source symbols in the decoded string?

1st three source symbols: "HHT"

4. [20 marks]
Consider four different *reduced* card decks in which each card has a *suit*, S (among *spade*, *heart*, *diamond*, and *club*), and a *rank*, R (among 1, 2, 3). For each of the four sub-decks in the columns of the table below, consider picking *one* card *at random* from the deck, and give the numeric values of the entropy-related quantities associated with that card and its suit and its rank.
(HINT: All 28 correct answers are among the 13 values at the bottom of the page.)

Quantity	Deck (i) (4 cards)	Deck (ii) (4 cards)	Deck (iii) (8 cards)	Deck (iv) (8 cards)
	spade 1 2 heart 1 2	spade 1 heart 1 2 3	spade 1 2 heart 1 2 diamond 1 2 club 1 2	spade 1 heart 1 2 diamond 1 2 club 1 2 3
$H(card) =$	2	2	3	3
$H(S) =$	1	.811	2	1.905
$H(R) =$	1	1.5	1	1.405
$H(R, S) =$	2	2	3	3
$H(S \mid R) =$	1	.5	2	1.594
$H(R \mid S) =$	1	1.189	1	1.094
$I(R, S) =$	0	.311	0	.311

Answers from which to choose:

$H(\frac{1}{2}, \frac{1}{2}) = 1.000$ $\quad \frac{1}{2}H(\frac{1}{2}, \frac{1}{2}) = 0.500$ $\quad = 0.000$ $\quad \frac{3}{2} - \frac{3}{4}\log_2(3) = 0.311$

$H(\frac{1}{4}, \frac{3}{4}) = 0.811$ $\quad \frac{3}{4}H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = 1.189$ $\quad \frac{1}{2}H(\frac{1}{2}, \frac{1}{2}) + \frac{3}{8}H(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = 1.094$

$H(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}) = 1.405$ $\quad H(\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{3}{8}) = 1.905$ $\quad \frac{1}{2}H(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) + \frac{3}{8}H(\frac{1}{3},\frac{1}{3},\frac{1}{3}) = 1.594$

$H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = 1.500$ $\quad H(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}) = 2.000$ $\quad H(\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8},\frac{1}{8}) = 3.000$

5. [20 marks]
For the source alphabet { a , b , c , d }, two different *first-order Markov models* are shown below, along with their corresponding *state-transition probabilities*. For each of the two, enter the values for the following quantities into the table below:

(i). What is the equilibrium state probability distribution, $\{T_a, T_b, T_c, T_d\}$?

(ii). What is the *entropy*, $H(X_n)$, of the state?

(iii). What is the *entropy rate*, $H(X_n \mid X_{n-1})$?

Model I

Transition Matrices:

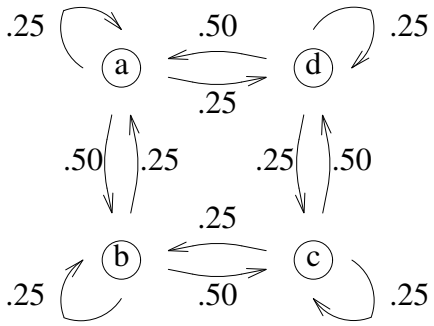
	a	b	c	d
a	¼	½	0	¼
b	¼	¼	½	0
c	0	¼	¼	½
d	½	0	¼	¼

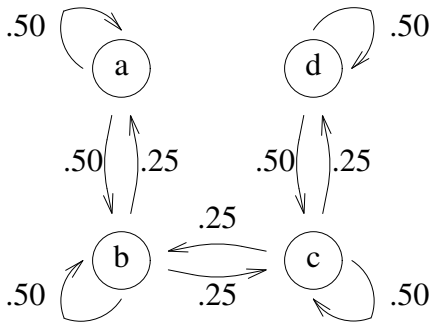
Model II

	a	b	c	d
a	½	½	0	0
b	¼	½	¼	0
c	0	¼	½	¼
d	0	0	½	½

[HINT: $\log_2(3) = 1.585$ **]**

	Model I	Model II
(i) $\{T_a, T_b, T_c, T_d\}$	$\{ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \}$	$\{ \frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6} \}$
(ii) $H(X_n)$	2	$\frac{1}{3} + \log_2(3) = 1.918$
(iii) $H(X_n \mid X_{n-1})$	1.5	$\frac{4}{3} = 1.333$


.25
.50
.25
a
d
.25
.50
.25
.25
.50
.25
b
c
.25
.50
.25


.50
.50
a
d
.50
.25
.50
.25
.25
b
c
.50
.25
.50