6.02 Recitation R1 (10 & 11 a.m.) 9/5/2013

- Self Intro: signal processing & speech
- T.A: Xue Feng
- Topics
  - review of probability
  - information, entropy, code
  - Huffman codes

Probability Model (three components associated w/ the model)

- sample space Ω: set of all possible outcomes ω of the experiment
- event: collection of subsets of the sample space
- probability measure:
  - $P(A) \geq 0$ where A is an event
  - $P(\Omega) = 1$
  - If $A \cap B = 0$, i.e., A & B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B) \qquad P(A \cup B) = P(A + B)$$

- Conditional probability: Prob of event A, given B has occurred

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) > 0$$

- Bayes' Rule: $P(B|A) = \dfrac{P(A|B)\,P(B)}{P(A)}$

- A & B are <u>independent</u> if

$$P(A|B) = P(A) \quad \text{or} \quad P(A \cap B) = P(A)\,P(B)$$

- <u>expected value</u> of random variable X

$$E[X] = \sum_x x\, p(x)$$

- <u>Definition of Information</u> (associated w) an event whose probability of occurrence is p

$$\boxed{I = \log_2\left(\frac{1}{p}\right) = -\log_2(p)}$$ (measured in binary info unit, or bit)

  - log to the base 2
  - some useful expressions

x	$\log_2 x$	
2	1	
3	1.58	
5	2.32	
7	2.80	
10	3.32	$(\log_2 2 + \log_2 5)$
100	6.64	

- information measures uncertainty:
  - higher uncertainty → lower probability → higher information
  - higher information is not a value statement; it is more surprising, (or it has a higher degree of suprises!)

- Entropy: expected value of the information obtained by learning the outcome of an event, S_i, $\{S_1, S_2, \cdots, S_i, \cdots, S_N\}$ w/ P_i $1 \le i \le N$

$$\boxed{H(P_1, P_2, \cdots, P_N) = \sum_{i=1}^{N} P_i I(P_i) = \sum_{i=1}^{N} P_i \log_2(1/P_i)}$$

- entropy is the average, or expected, uncertainty (amount of info) associated w/ the set of events

[Example]:

	Storrow Drive High Traffic volumn	Storrow Drive Low Traffic volumn
RedSox play	$p = 0.999$	$p = 0.001$
Red Sox $\overline{\text{play}}$	$p = 0.25$	$p = 0.75$

a) If RedSox are playing, how much info in bits is conveyed by the statement that the traffic volume is low?

$$I = \log_2(1/0.001) = \log_2(1000) = \log_2(10 \cdot 100) = 9.96 \text{ bits}$$

[note]: practically no info conveyed for high traffic volume when RedSox play

b) If Red Sox are not playing, what is the entropy of the corresponding probability distribution of traffic volume?

$$H = 0.25 \log_2(1/0.25) + 0.75 \log_2(1/0.75)$$

$$= 0.25 \log_2(4) + 0.75 (\log_2 4 - \log_2 3)$$

$$= 0.5 + 0.75 (2 - 1.58)$$

$$= 0.5 + 0.75 (0.42) = 0.815$$

- <u>Code</u> : mapping info we need to transmit into a bit sequence ; each bit sequence is called a codeword
- main objective of source coding is to compress the data by removing redundancy
  less time, less BW
- <u>fixed-length code</u> is inefficient (needs a lot of bits), but is easy to decode ; <u>variable-length code</u> is efficient, but decoding requires knowledge of prior history
- Given N symbols where symbol i occurs w/ prob p_i, then the <u>Expected Code Length</u> is given by

$$\boxed{L = \sum_i l_i p_i}$$

where l_i is the length of the i^{th} CW.
in bits

- entropy is the lower bound of the amount of info that must be sent, on average, when transmitting data. Shannon proved that:

$$\boxed{H \leq L}$$

- <u>Huffman Codes</u>
  - variable length code ; use shorter CW for more likely symbols & longer CW for less likely symbols
  - algorithm : see handout, reproduced from Lecture 1 (pg. 31)

[Example 2.7]

- messages made up entirely of the five vowel symbols. A, E, I, O, U

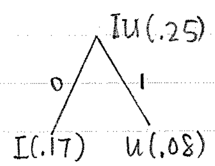
l	P_l	$\log_2(1/P_l)$	$P_l \log_2(1/P_l)$	CW	# of bits	codelength × P_l
A	.22	2.18	.48	10	2	.44
E	.34	1.55	.53	00	2	.68
I	.17	2.57	.43	010	3	.51
O	.19	2.40	.46	11	2	.38
U	.08	3.64	.29	011	3	.24
Total	1.0	12.34	2.19			Σ 2.25
			↑ entropy			↑ expected code length

a) info received (in bits) if known vowel is either "I" or "U"

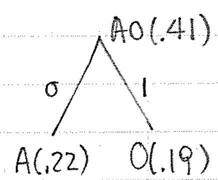
$$I = \log_2\left(\frac{1}{.17 + .08}\right) = \log_2 4 = 2 \text{ bits}$$

b) draw a Huffman Code Tree.

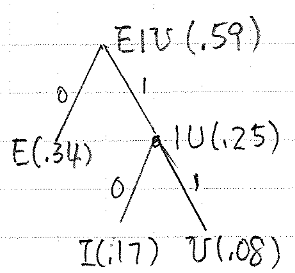
1)



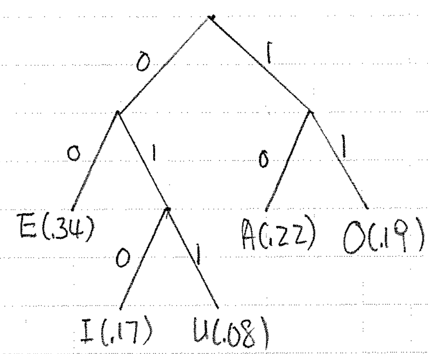
2)



3)



4)



c) expected length of transmitting 100 vowels = 100 × 2.25 = 225 bits.

d) Someone came up w/ expected length of 197 bits. Believabe? NO.

Because H = 2.19.

Observations on Huffman Codes

- non-unique : e.g., depends on how ties are broken, how "0" and "1" are assigned
- optimum : There are no other codes w/ shorter expected code length, when restricted to prefix-free code, and symbols that are drawn independently from known distributions
- It can be shown that the expected length satisfies

$$\boxed{H \leq L \leq H+1}$$

- instananeous code : The received CW can be decoded as soon as it is received
- prefix-free code : a code tree in which all symbols are at the leaves (prefix-free codes are naturally instantaneous)

Huffman's Coding Algorithm

- Begin with the set S of symbols to be encoded as binary strings, together with the probability p(s) for each symbol s in S.
- Repeat the following steps until there is only 1 symbol left in S:
  - Choose the two members of S having lowest probabilities. Choose arbitrarily to resolve ties.
  - Remove the selected symbols from S, and create a new node of the decoding tree whose children (sub-nodes) are the symbols you've removed. Label the left branch with a "0", and the right branch with a "1".
  - Add to S a new symbol that represents this new node. Assign this new symbol a probability equal to the sum of the probabilities of the two nodes it replaces.