

Solutions to Midterm Examination

1. (20 points) **Huffman coding.**

(a) (10 points) Find the binary Huffman code for the probability mass function

$$\mathbf{p} = \left(\frac{9}{39}, \frac{8}{39}, \frac{7}{39}, \frac{5}{39}, \frac{4}{39}, \frac{3}{39}, \frac{2}{39}, \frac{1}{39} \right).$$

(b) (10 points) Now find the 3-ary ($D = 3$) Huffman code for \mathbf{p} .

Solution to Huffman coding.

(a) One possible Huffman code is the following:

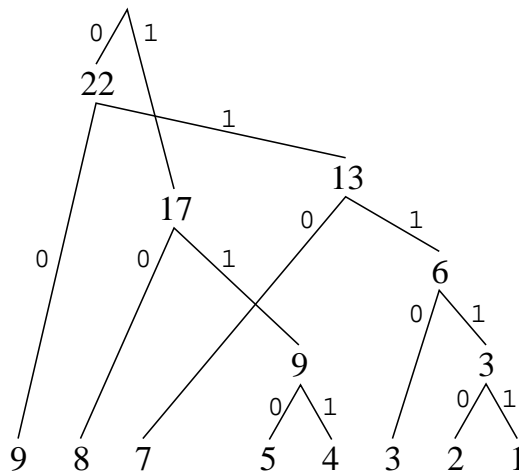


Figure 1: A Huffman tree for $D = 2$

The codewords in this case are:

symbol	probability	codeword	codeword length
x_1	9/39	00	2
x_2	8/39	10	2
x_3	7/39	010	3
x_4	5/39	110	3
x_5	4/39	111	3
x_6	3/39	0110	4
x_7	2/39	01110	5
x_8	1/39	01111	5

- (b) Since we have an even number of source symbols, we will need to add a dummy symbol to find the Huffman ternary code. One possible Huffman code is the following:

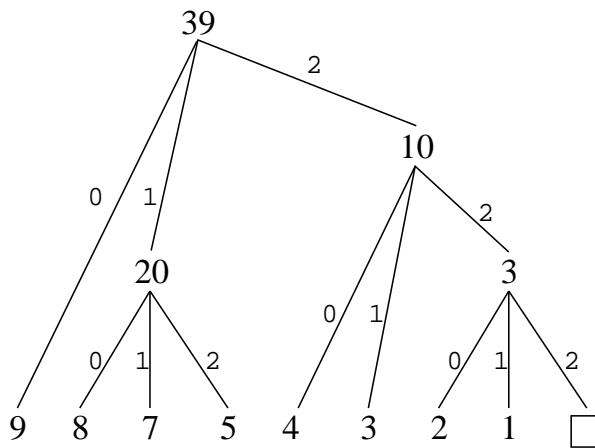


Figure 2: A Huffman tree for $D = 3$

The codewords in this case are:

symbol	probability	codeword	codeword length
x_1	9/39	0	1
x_2	8/39	10	2
x_3	7/39	11	2
x_4	5/39	12	2
x_5	4/39	20	2
x_6	3/39	21	2
x_7	2/39	220	3
x_8	1/39	221	3

2. (20 points) **Minimum expected description length code.**

Which of these codes cannot be optimal instantaneous codes for any \mathbf{p} :

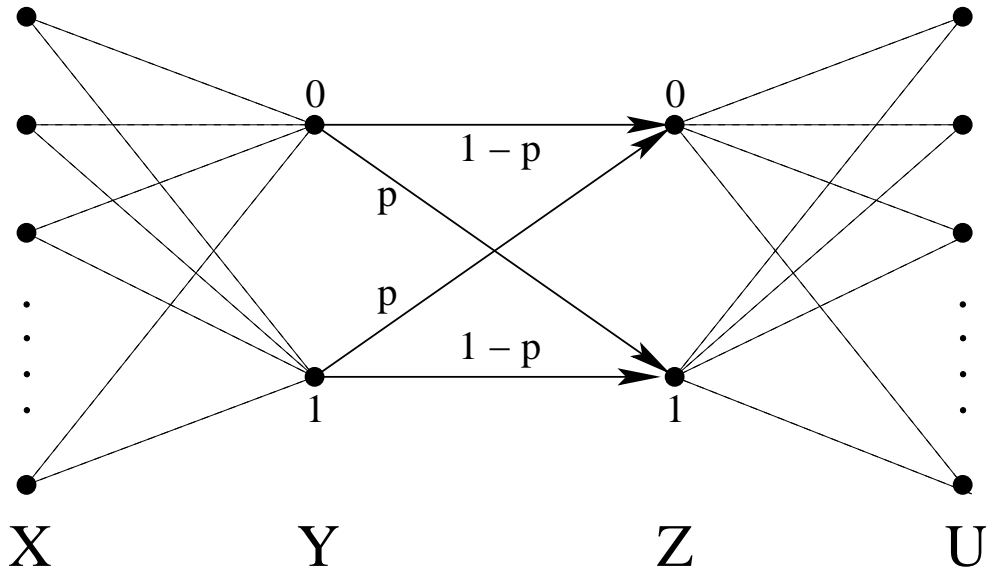
- (a) (10 points) 00,10,11
- (b) (10 points) 00,01,100,101,11

Solution to Minimum expected description length code.

- (a) This cannot be an optimal code since it can be trimmed down to $\{0, 10, 11\}$.
- (b) This is an optimal instantaneous code for $\{1/4, 1/4, 1/8, 1/8, 1/4\}$, respectively.

3. (30 points) Cascade.

Consider four random variables $X, Y, Z,$ and U . Let the alphabet sizes $|\mathcal{Y}| = |\mathcal{Z}| = 2$, and let the alphabet sizes $|\mathcal{X}|$ and $|\mathcal{U}|$ be arbitrary. Let $p(z|y)$ be as shown in the figure below. Let $p(y|x)$ and $p(u|z)$ be arbitrary.



Here $X \rightarrow Y \rightarrow Z \rightarrow U$ forms a Markov chain.

- (a) (20 points) Show $I(X; U) \leq 1 - H(p)$.
- (b) (10 points) What is a necessary condition on $p(y), y \in \{0, 1\}$, such that the bound $I(X; U) \leq 1 - H(p)$ can be achieved?

Solution to Cascade.

(a)

$$\begin{aligned} I(X; U) &\stackrel{(a)}{\leq} I(X; Z) \\ &\stackrel{(b)}{\leq} I(Y; Z) \\ &\stackrel{(c)}{=} H(Z) - H(Z|Y) \\ &\stackrel{(d)}{\leq} 1 - H(Z|Y) \\ &\stackrel{(e)}{=} 1 - \sum_{y \in \mathcal{Y}} p(Y = y)H(Z|Y = y) \\ &\stackrel{(f)}{=} 1 - (p(Y = 0)H(Z|Y = 0) + p(Y = 1)H(Z|Y = 1)) \\ &\stackrel{(g)}{=} 1 - (p(Y = 0)H(p) + p(Y = 1)H(p)) \\ &\stackrel{(h)}{=} 1 - H(p) \end{aligned}$$

where

- (a) data processing inequality
 - (b) data processing inequality, again
 - (c) definition of mutual information
 - (d) $H(Z) \leq \log_2 |\mathcal{X}| = \log_2 2 = 1$.
 - (e) definition of conditional entropy
 - (f) expanding the sum
 - (g) by the symmetry of the binary symmetric channel
 - (h) $p(Y = 0) + p(Y = 1) = 1$.
- (b) To achieve the upper bound, each of the inequalities must be satisfied with equality. The first two inequalities are beyond our control since $p(u|z)$ and $p(y|x)$ are unknown. However, we can provide a necessary condition for equality to hold in moving from step (c) to step (d) in the analysis above. In order for $H(Z) = 1$, Z should be Bernoulli($\frac{1}{2}$). Let $s = p(Y = 1)$. Then

$$\begin{aligned} p(Z = 0) &= (1 - s)(1 - p) + sp \\ p(Z = 1) &= (1 - s)p + s(1 - p). \end{aligned}$$

Setting $p(Z = 0) = p(Z = 1) = 1/2$, we have

$$\begin{aligned} (1 - s)p + s(1 - p) &= (1 - s)(1 - p) + sp \\ p - sp + s - sp &= 1 - s - p + sp + sp \\ 2s - 2sp - 2sp &= 1 - p - p \\ 2s(1 - 2p) &= 1 - 2p \\ s &= 1/2. \end{aligned}$$

Thus, a necessary condition for equality to hold is $p(Y = 1) = p(Y = 0) = 1/2$.

Some students fell into the pitfall of expanding $I(Y; Z)$ as $H(Y) - H(Y|Z)$, resulting in struggling more than necessary. Remember that there are multiple ways to expand, and since we are given the conditional distribution of Z given Y in this case, expanding $I(Y; Z)$ as $H(Z) - H(Z|Y)$ is a better idea.

4. (30 points) **Repeated histories.**

Let X_1, X_2, \dots be i.i.d. $\sim p(x)$.

(a) (20 points) Find the limiting behavior of

$$(p^r(X_1, X_2, \dots, X_n))^{\frac{1}{n}}$$

as $n \rightarrow \infty$.

(b) (10 points) Find the expected value

$$E \frac{1}{p(X_1, X_2, \dots, X_n)}$$

of $\frac{1}{p(X^n)}$.

Solution to Repeated histories.

(a)

$$\begin{aligned} (p^r(X_1, X_2, \dots, X_n))^{\frac{1}{n}} &= (p(X_1, X_2, \dots, X_n))^{\frac{r}{n}} \\ &= 2^{\log(p(X_1, X_2, \dots, X_n)) \frac{r}{n}} \\ &= 2^{\frac{r}{n} \log(p(X_1, X_2, \dots, X_n))} \\ &= 2^{r \frac{1}{n} \sum_{i=1}^n \log p(X_i)} \\ &\xrightarrow{a.s.} 2^{-rH} \end{aligned}$$

where the convergence occurs almost surely due to the Strong Law of Large Numbers.

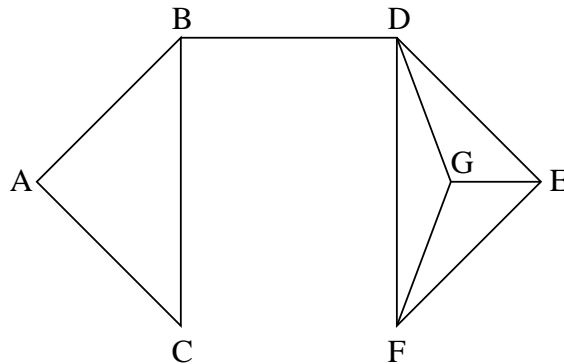
(b)

$$\begin{aligned} E \frac{1}{p(X_1, X_2, \dots, X_n)} &= E \frac{1}{p(X_1)p(X_2) \cdots p(X_n)} \\ &= E \frac{1}{p(X_1)} E \frac{1}{p(X_2)} \cdots E \frac{1}{p(X_n)} \\ &= \left(E \frac{1}{p(X)} \right)^n \\ &= \left(\sum_{x \in |\mathcal{X}|} p(x) \frac{1}{p(x)} \right)^n \\ &= \left(\sum_{x \in |\mathcal{X}|} 1 \right)^n \\ &= |\mathcal{X}|^n \end{aligned}$$

where the first two lines hold due to the independence of the X_i 's, and the third line holds due to the fact that the X_i 's are identically distributed.

A common mistake in this problem was to apply the AEP.

5. (70 points) **Random walks.**



Assume the stationary distribution on a random walk on this graph.

- (a) (10 points) Find $\Pr(\{X_1 = B\})$.
- (b) (10 points) Find $\Pr(\{X_1 X_2 X_3 X_4 X_5 = BABDE\})$.
- (c) (10 points) Find $\Pr(\{X_1 X_2 X_3 X_4 X_5 = EDBAB\})$.
- (d) (10 points) Is $\Pr(\{X^n = x_1 x_2 \dots x_n\})$ always equal to $\Pr(\{X^n = x_n x_{n-1} \dots x_1\})$?
Yes or No

You are offered 7-for-1 fair odds on gambles on the state of the walk. Assume log-optimal (maximal growth rate) gambling.

- (e) (10 points) What proportion of your wealth should you gamble on $X_1 = B$?
- (f) (10 points) Assuming a log-optimal betting scheme, in which each bet is based on the past states, and assuming $S_0 = 1$, how much wealth S_5 would you achieve on the sequence

$$X^5 = BABDE \text{ ?}$$

- (g) (10 points) Again assuming a log-optimal betting scheme, find $E \log S_n$.

Solution to Random walks.

- (a) The stationary distribution for a random walk on a connected undirected graph is given by the formula $\mu = \left(\frac{E_1}{2E}, \dots, \frac{E_n}{2E}\right)$, where E_i is the degree of node i and E is the total number of edges in the graph. Since the degree of node B is 3 and the total number of edges is 10, the stationary probability of being at node B is thus $\frac{3}{20}$.
- (b) By Markovity,

$$\begin{aligned} \Pr(\{X^5 = BABDE\}) &= \Pr(\{X_1 = B\}) \Pr(\{X_2 = A|X_1 = B\}) \cdots \\ &= \frac{3}{20} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{480}. \end{aligned}$$

- (c) As stated in the book, any random walk on a graph is time-reversible. Hence,

$$\Pr(\{X_1X_2X_3X_4X_5 = EDBAB\}) = \Pr(\{X_1X_2X_3X_4X_5 = BABDE\}) = \frac{1}{480}.$$

- (d) Yes, any random walk on a graph is time-reversible. It was not required to show a proof, but it is easy to give one. Note that by induction, it is sufficient to show that $\Pr(\{X_1X_2\}) = \Pr(\{X_2X_1\})$. That is, the Markov Chain must satisfy the detailed balance equations $\mu_i p_{ij} = \mu_j p_{ji}$:

$$\mu_i p_{ij} = \frac{E_i}{2E} \frac{1}{E_i} = \frac{1}{2E} = \frac{E_j}{2E} \frac{1}{E_j} = \mu_j p_{ji}.$$

- (e) Since $\Pr(\{X_1 = B\})$ is $\frac{3}{20}$, we should bet $\frac{3}{20} = 15\%$ of our wealth on this event.
- (f) Using proportional gambling, we would bet $\frac{1}{480}$ of our wealth on the event $X^5 = BABDE$. Since the odds are 7-to-1, our wealth would be $\frac{7^5}{480} \approx 35.015$.

(g) The general expression for the payoff after n moves is

$$S_n = 7^n b(X_1, X_2, \dots, X_n).$$

Since we use proportional gambling,

$$S_n = 7^n p(X_1, X_2, \dots, X_n).$$

Thus,

$$\begin{aligned} E \log S_n &= n \log 7 + E \log p(X_1, X_2, \dots, X_n) \\ &= n \log 7 - H(X_1, X_2, \dots, X_n) \\ &\stackrel{(a)}{=} n \log 7 - \left(\sum_{i=1}^n H(X_i | X_1^{i-1}) \right) \\ &\stackrel{(b)}{=} n \log 7 - \left(\sum_{i=1}^n H(X_i | X_{i-1}) \right) \\ &= n \log 7 - \left(H(X_1) + \sum_{i=2}^n H(X_i | X_{i-1}) \right) \\ &\stackrel{(c)}{=} n \log 7 - (H(X_1) + (n-1)H(X_2 | X_1)) \\ &\stackrel{(d)}{=} n \log 7 - (H(X_1) + (n-1)H(\mathcal{X})) \\ &\stackrel{(e)}{=} n \log 7 - (H(\mu) + (n-1)(\log(2E) - H(\mu))) \end{aligned}$$

where each step follows by

- (a) chain rule for entropy
- (b) Markovity
- (c) stationarity
- (d) for a stationary Markov chain, $H(\mathcal{X}) = H(X_2 | X_1)$
- (e) formula for entropy rate of random walk on a connected undirected graph.

Thus,

$$\begin{aligned} E \log S_n &= n \log 7 - (n-1) \log 2E + (n-2)H(\mu) \\ &= n(\log 7 - \log 2E + H(\mu)) + (\log 2E - 2H(\mu)) \end{aligned}$$

where

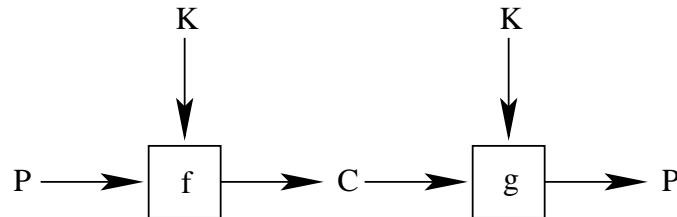
$$H(\mu) = H\left(\frac{E_1}{2E}, \dots, \frac{E_7}{2E}\right) = H\left(\frac{2}{20}, \frac{3}{20}, \frac{2}{20}, \frac{4}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}\right) \approx 2.771.$$

$E \log S_n$ grows linearly n with coefficient approximately equal to 1.256.

Few students received full credit for this problem. Common mistakes made included: 1) taking the limit as $n \rightarrow \infty$, although the problem asks for an answer for finite n , 2) using $H(\mu)$ where $H(\mathcal{X})$ should have been used, 3) allocating bets proportionally to the distribution μ in every time step, as if (X_i) were an i.i.d. sequence, and 4) giving an answer that did not scale with n .

6. (20 points) **Shannon cryptography.**

Consider the following relations between random variables P , K , and C :



- i. The plaintext P and the key K are independent of each other.
- ii. The ciphertext is a deterministic function of the plaintext and key: $C = f(P, K)$.
- iii. The plaintext is a deterministic function of the ciphertext and key: $P = g(C, K)$.

Show that the following relationship holds:

$$H(K|C) = H(K) + H(P) - H(C)$$

We conclude that if the ciphertext C gives no information about the key K , i.e., if $H(K|C) = H(K)$, then

$$H(C) = H(P).$$

Solution to Shannon cryptography.

As usual, there are many possible ways to do it. Here is a quick way:

$$\begin{aligned}
 H(K|C) &\stackrel{(a)}{=} H(K, C) - H(C) \\
 &\stackrel{(b)}{=} H(P, K, C) - H(C) \\
 &\stackrel{(c)}{=} H(K) + H(P|K) + H(C|P, K) - H(C) \\
 &\stackrel{(d)}{=} H(K) + H(P|K) - H(C) \\
 &\stackrel{(e)}{=} H(K) + H(P) - H(C)
 \end{aligned}$$

where each step follows by

- (a) chain rule,
- (b) P is a deterministic function of K and C ,
- (c) chain rule,
- (d) C is a deterministic function of P and K
- (e) by independence of P and K .

Hence, the uncertainty in the key given only the ciphertext is equal to the (unconditional) uncertainty of the key plus the uncertainty of the plaintext minus the uncertainty of the ciphertext.

Here is an alternative way using the familiar trick of expressing the same quantity in two different ways. Start with $H(P, K, C)$, which involves all the random variables under consideration. We now break this expression in two ways. The idea is to create either the term $H(C|P, K)$ or the term $H(P|C, K)$, since both of these are zero due to the latter two constraints mentioned in the problem.

$$\begin{aligned} H(P, K, C) &= H(P, K) + H(C|P, K) \\ &= H(P, K) \\ &= H(P) + H(K) \end{aligned}$$

where the last step holds due to the independence of P and K . Similarly,

$$\begin{aligned} H(P, K, C) &= H(C, K) + H(P|C, K) \\ &= H(C, K) \\ &= H(C) + H(K|C). \end{aligned}$$

Setting our two expressions for $H(P, K, C)$ equal to each other,

$$\begin{aligned} H(C) + H(K|C) &= H(P) + H(K) \\ H(K|C) &= H(P) + H(K) - H(C). \end{aligned}$$

There was a typo in the original writeup of the problem, in which it was incorrectly stated that f and g were one-to-one functions in two variables. Actually, we want them to be one-to-one functions in only one variable, when the key K is fixed; that is, $f(\cdot, K = k)$ is 1-1. This allows different ciphertexts to be created when the plaintexts are changed. Note that if f was 1-1 in two variables, one could deduce both the key and plaintext from the ciphertext alone, which undermines the point of having a cryptosystem. However, the few students who cleverly exploited this typo to solve the problem were still given full credit.