

# 520.447: Information Theory and Coding

## Final Examination

2:00 — 5:00 PM, May 5, 2000.

Read these instructions before starting the examination.

- (i) This is a closed-book examination. Use of the textbook, notes, *etc.*, is not permitted. Some useful formulae appear at the end of the examination booklet.
- (ii) Use of electronic calculators is permitted for numeric calculations only.
- (iii) Show all your work clearly and concisely. Points may be deducted for illegible or unclear answers.
- (iv) Write your answers in the space provided. Use the unprinted side of the pages in the examination booklet if necessary.
- (v) There are six mandatory questions for a total of 100 points. You must answer **all** parts of these questions. There is an optional *bonus question*. Points earned on the bonus question will be added to your total. Students enrolled in the Ph.D. program are *strongly encouraged* to attempt the bonus question.
- (vi) If you want your exam score and final grade sent to you by electronic mail, please enter your e-mail address in the space provided below. If you do not wish your grade to be sent via electronic mail, please write “do not post grade” in this space.

Best of luck!

---

Name: \_\_\_\_\_

E-mail: \_\_\_\_\_

---

Question N <sub>0</sub> 1	/10 Points
Question N <sub>0</sub> 2	/20 Points
Question N <sub>0</sub> 3	/20 Points
Question N <sub>0</sub> 4	/20 Points
Question N <sub>0</sub> 5	/20 Points
Question N <sub>0</sub> 6	/10 Points
Bonus Question	/20 Points

---

TOTAL /100 Points

---

**Question No 1:** *Entropy of a Disjoint Mixture.* Let  $X_1$  and  $X_2$  be discrete random variables drawn according to probability mass functions  $p_1(\cdot)$  and  $p_2(\cdot)$  over the respective alphabets  $\mathcal{X}_1 = \{1, 2, \dots, m\}$  and  $\mathcal{X}_2 = \{m + 1, m + 2, \dots, n\}$ . Let

$$X = \begin{cases} X_1 & \text{with probability } \alpha, \text{ and} \\ X_2 & \text{with probability } 1 - \alpha. \end{cases}$$

1(a) Find  $H(X)$  in terms of  $H(X_1)$ ,  $H(X_2)$  and  $\alpha$ . (4 points)

1(b) Maximize  $H(X)$  over  $\alpha$  to show that  $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ , and interpret the result using the notion that  $2^{H(X)}$  is the effective alphabet size. (6 points)

**Question No 2:** Let  $P$  be a probability mass function on a discrete set  $\mathcal{X}$ , and  $c : \mathcal{X} \rightarrow \mathcal{D}^*$  be a uniquely decodable  $D$ -ary code for  $\mathcal{X}$ . Let  $\ell(x) = |c(x)|$  denote the length of the codeword  $c(x)$  for a symbol  $x \in \mathcal{X}$ .

2(a) Shannon's first coding theorem states that the expected or average codeword length for *any* uniquely decodable code for  $\mathcal{X}$  cannot be any smaller than the entropy  $H(P)$ .

$$\begin{aligned}
 E[\ell(X)] - H(X) &= \sum_{x \in \mathcal{X}} P(x)\ell(x) + \sum_{x \in \mathcal{X}} P(x) \log_D P(x) \\
 &= \sum_{x \in \mathcal{X}} P(x) \left[ -\log_D D^{-\ell(x)} \right] + \sum_{x \in \mathcal{X}} P(x) \log_D P(x) \\
 &= \sum_{x \in \mathcal{X}} P(x) \log_D \left( \frac{P(x)}{D^{-\ell(x)}} \right) \\
 &= \sum_{x \in \mathcal{X}} P(x) \log_D \left( \frac{P(x)}{\sum_{x' \in \mathcal{X}} D^{-\ell(x')} \cdot \sum_{x' \in \mathcal{X}} D^{-\ell(x')}} \right) \\
 &\quad \vdots \\
 &\geq 0
 \end{aligned}$$

Complete the missing steps in this proof; justify your steps.

(10 points)

2(b) For  $\mathcal{X} = \{a, b, c, d, e, f, g, h\}$  and  $P = \left(\frac{1}{3}, \frac{3}{18}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}\right)$ , design a *ternary prefix-free code*  $c : \mathcal{X} \rightarrow \{0, 1, 2\}^*$  with the shortest possible average codeword length. Write the final codeword assignment in the table below. Compute the average codeword length for this code and compare it with the entropy of  $P$ . (10 points)

$x$	$c(x)$
$a$	
$b$	
$c$	
$d$	
$e$	
$f$	
$g$	
$h$	

$$E[\ell(X)] =$$

$$H(P) =$$

**Question No 3:** In homework assignment no 12, we computed the information capacity of a *sum channel* with two subchannels  $W_1$  and  $W_2$  to be

$$C = \log_2 \{2^{C_1} + 2^{C_2}\}, \quad (1)$$

where  $C_1$  is the capacity of  $W_1$  and  $C_2$  of  $W_2$ . Now, let  $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3$  and  $\mathcal{Y} = \mathcal{Y}_1 \cup \mathcal{Y}_2 \cup \mathcal{Y}_3$  represent the input and output alphabet respectively of a discrete memoryless sum channel with *three* subchannels. Recall that in a sum channel, there is no cross-over from input symbols in  $\mathcal{X}_i$  to output symbols in  $\mathcal{Y}_j$  for  $i \neq j$ . Stated differently, the channel transition probabilities can be arranged as a block diagonal matrix:

$$[p(y|x)] = [w_{xy}] = W = \begin{bmatrix} W_1 & 0 & 0 \\ 0 & W_2 & 0 \\ 0 & 0 & W_3 \end{bmatrix} \quad (2)$$

3(a) Draw the channel transition diagram for the specific case of

$$\mathcal{X}_1 = \{1, 2\} = \mathcal{Y}_1, \quad \mathcal{X}_2 = \{3, 4\} = \mathcal{Y}_2, \quad \mathcal{X}_3 = \{5, 6\} = \mathcal{Y}_3,$$

where each subchannel  $W_1$ ,  $W_2$  and  $W_3$  is a binary symmetric channel with cross-over probability  $\frac{1}{2}$ . Write the  $6 \times 6$  transition matrix  $W$  for this channel. (2 points)

3(b) What is the capacity of the channel described in 3(a) above? (2 points)

- 3(c) Calculate the capacity  $C$  of the general sum channel  $W$  of equation (2), *not* the specific case of 3(a), in terms of the capacities  $C_1$  of  $W_1$ ,  $C_2$  of  $W_2$  and  $C_3$  of  $W_3$ . [*Hint*: The short way to solve this problem is to carefully extend the result (1) of two subchannels. The slightly longer but equally correct way is to define a random variable

$$Z = \begin{cases} 1 & \text{if } X \in \mathcal{X}_1, \\ 2 & \text{if } X \in \mathcal{X}_2, \\ 3 & \text{if } X \in \mathcal{X}_3, \end{cases}$$

and exploit the equalities  $H(Y) = H(Y, Z)$  and  $H(Y|X) = H(Y|X, Z)$  (why?) to express  $I(X; Y)$  in terms of  $I(X; Y|Z)$  and  $H(Z)$ .] (8 points)

- 3(d) Compute the capacity of a channel with input alphabet  $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , output alphabet  $\mathcal{Y} = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ , and channel transition probabilities  $p(y|x)$  given by

$$[p(y|x)] = [w_{xy}] = \begin{bmatrix} \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \end{bmatrix}.$$

[*Hint*: Draw the channel transition diagram.]

(8 points)

## Extra Work Space

**Question No 4:** *The two-look Gaussian channel.* Consider the usual discrete time memoryless Gaussian channel without feedback, with the modification that the transmitted symbol  $X$  is received at two distinct receivers as  $Y_1$  and  $Y_2$ , and the decoder has access to both  $Y_1$  and  $Y_2$ . This, for example, is the situation in an antenna-array. The noise in the two observations is, in general, going to be correlated. Assume that

$$\begin{aligned} Y_1 &= X + Z_1, \\ Y_2 &= X + Z_2, \end{aligned}$$

the transmitter has a signal power constraint  $P$ , and the noise vector  $Z = [Z_1 \ Z_2]^T$  is independent of  $X$  and is jointly Gaussian with zero mean and covariance matrix

$$K_Z = \begin{bmatrix} N & \rho N \\ \rho N & N \end{bmatrix}$$

4(a) Compute the capacity of this channel in terms of  $P$ ,  $N$  and  $\rho$ . (10 points)

4(b) What is the capacity when  $\rho = 1$ ? Interpret your answer by comparing it with the capacity of a *one observation* Gaussian channel with power constraint  $P$ . (3 points)

4(c) What is the capacity when  $\rho = 0$ ? Interpret your answer by comparing it with the capacity of a *one observation* Gaussian channel with power constraint  $P$ . (3 points)

4(d) What is the capacity when  $\rho = -1$ ? Provide an intuitive explanation for the answer by describing a simple decoding scheme which achieves capacity. (4 points)

**Question No 5:** Let  $P$  be a probability mass function on a discrete alphabet  $\mathcal{X}$ , and let  $d_0 : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}$  be a bounded distortion measure for quantizing symbols from  $\mathcal{X}$  into symbols from  $\hat{\mathcal{X}}$ . Let  $|\mathcal{X}| = N$  and  $|\hat{\mathcal{X}}| = N'$ . Note that  $d_0$  can always be specified as a  $N \times N'$  matrix

$$d_0(x, \hat{x}) = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,N'} \\ d_{2,1} & d_{2,2} & \cdots & d_{2,N'} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N,1} & d_{N,2} & \cdots & d_{N,N'} \end{bmatrix}.$$

Let the encoder-decoder pair  $f : \mathcal{X} \rightarrow \{1, \dots, M\}$  and  $g : \{1, \dots, M\} \rightarrow \hat{\mathcal{X}}$  of a rate  $R = \log M$  quantizer be compactly written as  $q(x) = g(f(x))$ .

5(a) For any constant  $\alpha$ , define a new distortion measure

$$d_\alpha(x, \hat{x}) = d_0(x, \hat{x}) - \alpha.$$

Show that the average distortion of a quantizer under the two different distortion measures is related as (3 points)

$$E[d_0(X, q(X))] = E[d_\alpha(X, q(X))] + \alpha.$$

This relationship implies that if a sequence of rate- $R$  quantizers achieves distortion  $D$  under  $d_\alpha$ , then that *same* sequence of quantizers achieves distortion  $D + \alpha$  under  $d_0$ .

5(b) Justify the assumption often made in calculating the rate distortion function that the distortion measure is nonnegative. In other words show how the rate distortion function  $R_0(D)$  for a distortion measure  $d_0$  which *does not satisfy the assumption* can be easily obtained from the rate distortion function  $R_\alpha(D)$  for a different distortion measure  $d_\alpha$

which is nonnegative. [Hint: Choose a good  $\alpha$ .]

(3 points)

5(c) For a collection of constants  $\alpha(x)$ ,  $x \in \mathcal{X}$ , define a new distortion measure

$$d_{\underline{\alpha}}(x, \hat{x}) = d_0(x, \hat{x}) - \alpha(x). \quad (3)$$

Show that the average distortion of a quantizer under the two different distortion measures is related as (3 points)

$$E[d_0(X, q(X))] = E[d_{\underline{\alpha}}(X, q(X))] + E[\alpha(X)].$$

Note that since  $\hat{\alpha} = E[\alpha(X)]$  does not depend on the quantizer, this relationship implies that if a sequence of rate- $R$  quantizers achieves distortion  $D$  under  $d_{\underline{\alpha}}$ , then that *same* sequence of quantizers achieves distortion  $D + \hat{\alpha}$  under  $d_0$ .

- 5(d) Justify the assumption often made in calculating the rate distortion function that for every input symbol, there is at least one reproduction symbol with zero distortion. In other words show how the rate distortion function  $R_0(D)$  for a distortion measure  $d_0$  which *does not* satisfy the assumption can be easily obtained from the rate distortion function  $R_{\underline{\alpha}}(D)$  for a different distortion measure  $d_{\underline{\alpha}}$  which *does* satisfy the assumption. [Hint: Choose a good  $\underline{\alpha}$ .] (3 points)

- 5(e) Let  $\mathcal{X} = \{0, 1\} = \hat{\mathcal{X}}$ , and let the distortion measure be

$$d_0(x, \hat{x}) = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}. \quad (4)$$

Clearly state the range  $D_{\min} < D < D_{\max}$  in which  $0 < R_0(D) < 1$  for a Bernoulli random variable with  $P(0) = P(1) = \frac{1}{2}$ . [Hint: Try  $\alpha(0) = 1$  and  $\alpha(1) = 2$  in equation (3).] (3 points)

5(f) Calculate the rate distortion function  $R_0(D)$  for a Bernoulli random variable with  $P(0) = P(1) = \frac{1}{2}$  for the distortion measure  $d_0$  of equation (4). (5 points)

**Question No 6:** Consider an urn containing 50 red, 30 blue and 20 green balls. Let  $X_1, X_2, \dots, X_n$  be a random sequence of  $n$  balls drawn from the urn *with replacement*, where  $X_i \in \{\text{red, blue, green}\}$ . Let  $R_n = R(X_1, X_2, \dots, X_n)$  be the number of red balls,  $B_n = B(X_1, X_2, \dots, X_n)$  the number of blue balls, and  $G_n = G(X_1, X_2, \dots, X_n)$  the number of green balls in the sequence.

- 6(a) Write an expression for the probability of a sequence with  $R_n = r$ ,  $B_n = b$ , and  $G_n = g$ . Evaluate this expression for  $R_n = 5$ ,  $B_n = 3$ , and  $G_n = 2$ . (3 points)

- 6(b) State information theoretic upper and lower bounds for the number of sequences with  $R_n = r$ ,  $B_n = b$ , and  $G_n = g$ . Calculate the exact number of sequences for  $R_n = 5$ ,  $B_n = 3$ , and  $G_n = 2$ . Compare this number with your bounds. (4 points)

- 6(c) State information theoretic upper and lower bounds for the *total* probability of *all sequences* with  $R_n = r$ ,  $B_n = b$ , and  $G_n = g$ . Calculate the exact probability for  $R_n = 5$ ,  $B_n = 3$ , and  $G_n = 2$ . Compare this probability with your bounds. (3 points)

**Bonus Question:** This problem develops another lower bound, similar to the *Hamming* bound of homework assignment no 12, on the error correcting capability of a  $(n, k)$  *binary* linear block code in terms of  $n$  and  $k$ . The bound developed here is called the *Plotkin* bound.

- 7(a) Show that minimum Hamming distance  $d_{\min}$  between two codewords of a binary linear block code is equal to the Hamming weight of the codeword with the smallest number of 1's, excluding the all-0 codeword. (4 points)

Consider an  $(n, k)$  binary linear block code whose generator matrix  $\mathbf{G}$  does not contain any all-zero row. An example of such a matrix is

$$\mathbf{G} = \begin{bmatrix} I_{4 \times 4} \\ P_{3 \times 4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}, \quad (5)$$

the generator of the  $(7, 4)$  Hamming code we studied in class. A  $k$ -bit message vector  $\mathbf{m}$  generates a  $n$ -bit codeword vector  $\mathbf{c} = \mathbf{G} \times \mathbf{m}$ . *e.g.* For

$$\mathbf{m} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Let  $\mathbf{C}$  be the  $n \times 2^k$  matrix in which each column is an  $n$ -bit codeword for one of the  $2^k$  distinct  $k$ -bit messages. For the generator matrix of equation (5), for example,

$$\mathbf{C} = \mathbf{G} \times \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

- 7(b) Show that for any  $(n, k)$  binary linear block code, each row of  $\mathbf{C}$  contains exactly  $2^{k-1}$  0's and  $2^{k-1}$  1's. [*Hint*: What is the relationship between the  $(i, j)$ -th component of  $\mathbf{C}$  and the  $i$ -th row of  $\mathbf{G}$ ? What happens as  $j$  goes from 0 to  $2^k$ ?] (8 points)

7(c) Show that the minimum distance  $d_{\min}$  of an  $(n, k)$  binary linear block code satisfies

$$d_{\min} \leq \frac{n \cdot 2^{k-1}}{2^k - 1}. \quad (6)$$

[*Hint:* The result of 7(b) states how many 1's there are in the ensemble  $\mathbf{C}$  of all the codewords. The result of 7(a) states how  $d_{\min}$  relates to the number of 1's in one particular codeword.] (8 points)

## Extra Work Space