16:332:542 **Information Theory and Coding** **Final Examination**

May 5, 2005

This is an 180 minute exam. Please answer the following questions in the notebooks provided. This is a closed book test. Make sure that you have included your name, personal 4 digit code (unrelated to your RU ID digits) and signature in each book used (*5 points*). Read each question carefully. All statements must be justified. Computations should be simplified as much as possible.

1. *20 points* Let $X_1, Z_1, Z_2, \ldots$ be iid Bernoulli random variables which take values 0 and 1 with equal probability. Define the sequence of random variables X_i as

$$X_{i+1} = X_i + Z_i, \qquad i = 1, 2, \ldots, n-1.$$

Find the mutual information $I(X_1; X_2, X_3, \ldots, X_n)$.

From the definition of mutual information,

$$\begin{aligned} I(X_1; X_2, X_3, \ldots, X_n) &= H(X_2, \ldots, X_n) - H(X_2, \ldots, X_n | X_1) \\ &= H(X_2) + \sum_{i=3}^{n} H(X_i | X_{i-1}, \ldots, X_2) - \sum_{i=2}^{n} H(X_i | X_{i-1}, \ldots, X_1) \end{aligned}$$

Note that for $1 \le j \le i-1$,

$$H(X_i | X_{i-1}, \ldots, X_j) = H(X_{i-1} + Z_{i-1} | X_{i-1}, \ldots, X_j) = H(X_{i-1} + Z_{i-1} | X_{i-1}) = H(Z_{i-1})$$

since Z_{i-1} is independent of the prior $X_{i-1}, X_{i-2}, \ldots, X_j$. Thus

$$\begin{aligned} I(X_1; X_2, X_3, \ldots, X_n) &= H(X_2) + \sum_{i=3}^{n} H(Z_{i-1}) - \sum_{i=2}^{n} H(Z_{i-1}) \\ &= H(X_2) - H(Z_1) \end{aligned}$$

In addition, the PMF of $X_2 = X_1 + Z_1$ is

$$P_{X_2}(x) = \begin{cases} 1/4 & x = 0, 2 \\ 1/2 & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

It follows that

$$H(X_2) = -2\left[\frac{1}{4}\log\frac{1}{4}\right] - \frac{1}{2}\log\frac{1}{2} = \frac{3}{2}.$$

Since $H(Z_1) = 1$, $I(X_1; X_2, X_3, \ldots, X_n) = 1/2$.

---

2. *35 points* Let Z take values 0 and 1 with probabilities $1-p$ and p. Let X, which is independent of Z, take values $1, 2, \ldots, n$ with probabilities $\mathbf{q} = [q_1, q_2, \ldots, q_n]$. Let

$$Y = XZ.$$

(a) *10 points* Find the entropy of Y in terms of the entropies of X and Z.

The PMF of Y is

$$P_Y(y) = \begin{cases} 1-p & y=0 \\ pq_i & y=1,2,\ldots,n \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the entropy of Y is

$$\begin{aligned} H(Y) &= -(1-p)\log(1-p) - \sum_{i=1}^{n} pq_i \log(pq_i) \\ &= -(1-p)\log(1-p) - p\sum_{i=1}^{n} q_i (\log p + \log q_i) \\ &= -(1-p)\log(1-p) - p\log p \underbrace{\sum_{i=1}^{n} q_i}_{1} - p\sum_{i=1}^{n} q_i \log q_i \\ &= H(Z) + pH(X) \end{aligned}$$

(b) *10 points* Find the p and $\mathbf{q}$ that maximize $H(Y)$.

For any p, we want to choose $\mathbf{q}$ to maaximize $H(X)$. This is done by choosing $\mathbf{q} = [1/n, \ldots, 1/n]$, which yields $H(X) = \log n$. With this choice of $\mathbf{q}$,

$$H(Y) = H(p) + p\log n.$$

Working in nats, we find the optimal p via

$$\frac{dH(Y)}{dp} = \log(1-p) - \log p + \log n = 0.$$

This implies $H(Y)$ is maximized at $p = n/(n+1)$. In fact, for this choice of p, all values of Y are equiprobable and $H(Y) = \log(n+1)$.

(c) *15 points* Suppose X and Y are the input and output of a discrete memoryless channel. For fixed p, what is the capacity $C(p)$ of the channel? What value of p maximizes $C(p)$?

First we find the mutual information

$$I(X;Y) = H(Y) - H(Y|X).$$

Fortunately, we know already know $H(Y) = H(Z) + pH(X)$. This leaves

$$H(Y|X) = H(XZ|X) = H(Z).$$

Thus $I(X;Y) = pH(X)$. This is not surprising $Z = 0$ erases the symbol X. Essentially the channel is an n input erasure channel.

For fixed $p > 0$, $I(X;Y) = pH(X)$ is maximized by choosing $\mathbf{q} = [1/n, \ldots, 1/n]$ so as to maximize $H(X)$. Thus $C(p) = p\log n$. Finally, $C(p)$ is maximized at $p = 1$. That is, the capcity of the erasure channel is highest when there are no erasures.

3. *40 points* A discrete memoryless multiple access channel has inputs X_1 and X_2 and output $Y = X_1 + X_2$. The inputs X_1 and X_2 both use alphabet $\mathcal{X} = \{0,1,2\}$; the output Y has alphabet $\mathcal{Y} = \{0,1,\ldots,4\}$.

(a) *20 points* Under the assumption that each X_i uses equiprobable inputs, find and sketch the (R_1, R_2) region of achievable rates for this 2-user MAC.

Under the assumption that

$$P_{X_i}(x) = \begin{cases} 1/3 & x = 0,1,2, \\ 0 & \text{otherwise,} \end{cases}$$

The achievable rate region of the 2-user MAC is given by the constraints

$$R_1 \le I(X_1; Y|X_2), \qquad R_2 \le I(X_2; Y|X_1), \qquad R_1 + R_2 \le I(X_1, X_2; Y).$$

Note that

$$\begin{aligned} I(X_1; Y|X_2) = I(X_1; X_1 + X_2|X_2) &= H(X_1|X_2) - H(X_1|X_1 + X_2, X_2) \\ &= H(X_1|X_2) - H(X_1|X_1, X_2) \\ &= H(X_1|X_2) = H(X_1) \end{aligned}$$

By symmetry, or by a symmetric sequence of steps if you don't believe in symmetry, we can conclude that $I(X_2; Y|X_1) = H(X_2)$.
For the sum rate constraint,

$$I(X_1, X_2; Y) = H(Y) - H(Y|X_1, X_2) = H(Y)$$

since $Y = X_1 + X_2$ implies $H(Y|X_1, X_2) = 0$. The PMF of Y is the convolution of the PMFs of X_1 and X_2. From the following table of Y as a function of X_1 and X_2,

Y	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$X_1 = 0$	0	1	2
$X_1 = 1$	1	2	3
$X_1 = 2$	2	3	4

Since each X_1, X_2 pair has probability $1/9$,

$$P_Y(y) = \begin{cases} 1/9 & y = 0,4, \\ 2/9 & y = 1,3, \\ 3/9 & y = 2, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned} H(Y) &= -2\left[\frac{1}{9}\log\frac{1}{9}\right] - 2\left[\frac{2}{9}\log\frac{2}{9}\right] - \frac{3}{9}\log\frac{3}{9} \\ &= \frac{15\log 3 - 4}{9} = 2.197 \text{ bits} \end{aligned}$$

Since $H(X_1) = H(X_2) = \log 3 = 1.585$ bits, the rate region (which all of you should have skteched) is

$$R_1 \le 1.585 \qquad R_2 \le 1.585, \qquad R_1 + R_2 \le 2.197.$$

(b) *20 points* Suppose user 1 and user 2 collaborate and act as single transmitter of rate R with input $\mathbf{X} = (X_1, X_2)$ and output Y. What is the capacity of the channel? What input distribution achieves capacity?

Since $Y = X_1 + X_2$, $H(Y|\mathbf{X}) = 0$ and

$$I(\mathbf{X}; Y) = H(Y) - H(Y|\mathbf{X}) = H(Y).$$

Thus, we achieve capacity by maximizing the output entropy $H(Y)$. Since $Y \in \{0, 1, \ldots, 4\}$, we know that $H(Y) \le \log 5$. This upper bound is achieved if we can make the outputs equiprobable. referring to the table for the value of Y as a function of X_1 and X_2 in the previous part, we see that this is possible in many ways. One such way is the following joint PMF

$P_{X_1,X_2}(x_1, x_2)$	$x_2 = 0$	$x_2 = 1$	$x_2 = 2$
$X_1 = 0$	$1/5$	$1/10$	$1/15$
$X_1 = 1$	$1/10$	$1/15$	$1/10$
$X_1 = 2$	$1/15$	$1/10$	$1/5$

Any other joint PMF such that each anti-diagonal sums to $1/5$ will also achieve capacity. Also, note that $\log 5 = 2.322 > 2.197$. That is, a cooperative strategy achieves a sum rate that is about 0.1 bits per channel use higher than that achieved by independent signaling.

---

4. *20 points* Consider two parallel channels with independent Gaussian noise Z_1 and Z_2 with variances $N_1 = 1$ and $N_2 = 2$. The signalling is

$$Y_1 = X_1 + Z_1$$
$$Y_2 = X_2 + Z_2$$

The transmitter is subject to the power constraint $E\left[X_1^2 + X_2^2\right] \le \overline{P}$. Find and sketch the capacity $C(\overline{P})$ of this channel as a function of $\overline{P}$.

This problem is a gift. It was going to be harder ... but it seemed like the exam was going to be too long. The solution, of course, is the waterfilling allocation. We choose powers $P_i = E\left[X_i^2\right]$ such that

$$P_i = (\lambda - N_i)^+.$$

where λ is chosen so that $P_1 + P_2 = \overline{P}$. In this problem, $N_1 = 1$ and $N_2 = 2$, so that

$$P_1 = (\lambda - 1)^+, \qquad P_2 = (\lambda - 2)^+$$

Since the channels are Gaussian,

$$\begin{aligned} C(\overline{P}) &= \frac{1}{2} \sum_i \log\left(1 + \frac{P_i}{N_i}\right) \\ &= \frac{1}{2} \log(1 + (\lambda - 1)^+) + \frac{1}{2} \log\left(1 + \frac{(\lambda - 2)^+}{2}\right). \end{aligned}$$

The average power constraint implies

$$(\lambda - 1)^+ + (\lambda - 2)^+ = \overline{P}.$$

For very small $\overline{P}$, we obtain $1 \leq \lambda < 2$, implying $(\lambda - 2)^+ = 0$. It follows that $\lambda = 1 + \overline{P}$ and that

$$C(\overline{P}) = \frac{1}{2}\log(1+\overline{P}), \qquad \overline{P} \text{ small}$$

---

5. *55 points* Every coding theorem we proved this semester included a converse that was proven using the Fano bound. For example, in the case of a discrete, memoryless channel, for any sequence of $(2^{nR}, n)$ codes with message index X, codewords $X^n(W)$, receiver output Y^n, decoding function $g(Y^n)$, and error probability $P_e^{(n)} = P\left[W \neq g(Y^n)\right]$, the proof used these steps:

$$\begin{aligned} nR &= H(W) && (1)\\ &= H(W|Y^n) + I(W;Y^n) && (2)\\ &\leq H(W|Y^n) + I(X^n(W);Y^n) && (3)\\ &\leq 1 + P_e^{(n)}nR + I(X^n(W);Y^n) && (4)\\ &\leq 1 + P_e^{(n)}nR + nC && (5) \end{aligned}$$

   (a) *25 points* For each of the above steps, (1) through (5), there is a specific reason that step holds. Given a precise justification for each of the five steps above.

   (b) *10 points* Explain how step (5) implies a converse to the coding theorem.

   (c) *20 points* For one of the above five steps, the correct reason is simply "the Fano bound" or "Fano's inequality." *Derive the Fano bound as used in the above five step proof.* Hint: The proof defines the error event

$$E = \begin{cases} 1 & g(Y^n) \neq W \\ 0 & g(Y^n) = W \end{cases}$$

   and then expands $H(E, W|Y^n)$ in two different ways.