ECE1502F — Information Theory
Final Examination — Solutions
December 11, 2007

1. (a) Let X and Y be independent Bernoulli(1/2) random variables and let $Z = X \oplus Y$, where $\oplus$ denotes modulo-two addition. Since X and Y are independent, we have $I(X;Y) = 0$. Since X is equally likely to be zero or one for each possible value of Z, Z provides no information about X, i.e., $I(X;Z) = 0$. However, $I(X;Y,Z) = 1$, since knowledge of both Y and Z determines X uniquely. Thus the answer to the question is **no**.

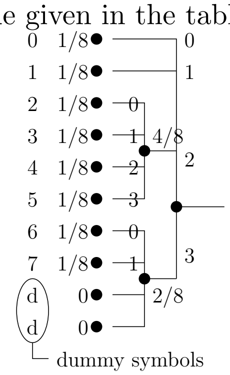
   (b) The same example as in the previous question provides a counterexample, i.e, $I(X;Y) = 0$, yet $I(X;Y|Z) = 1$.

   (c) **True**. To see this, let X be a postive real-valued random variable over with $E[X] = m$ and probability density function $p(x)$. and let Y be an exponential random variable with probability density function $f(x) = (1/m)e^{-x/m}$. Then

$$
\begin{aligned}
0 &\leq D(p||f) \\
&= \int_0^\infty p(x)\ln[p(x)/f(x)]dx \\
&= \int_0^\infty p(x)\ln p(x)dx + \int_0^\infty p(x)\ln(1/f(x))dx \\
&= -h(X) + \int_0^\infty p(x)\ln(me^{x/m})dx \\
&= -h(X) + \frac{1}{m}\int_0^\infty xp(x)dx + \ln(m)\int_0^\infty f(x)dx \\
&= -h(X) + 1 + \ln(m) \\
&= -h(X) + \ln(me).
\end{aligned}
$$

   Thus we find that $h(X) \leq \ln(me)$ with equality achieved if and only if $p(x) = f(x)$ a.e., i.e., if and only if X is an exponential random variable.

   (d) To design a Huffman code over a quaternary alphabet it is necessary that the number of probability masses be one larger than an integer multiple of 3. Thus, we must add two dummy symbols. The resulting tree is shown in the figure, and corresponds to the code given in the table.



x	0	1	2	3	4	5	6	7
$C(x)$	0	1	20	21	22	23	30	31

   (e) The channel is symmetric, and so has capacity $C = \log_2(3) - H(1/2, 1/4, 1/4) = \log_2(3) - 3/2 \approx 0.085$ bits per channel use.

2. (a) The channel is neither symmetric nor weakly symmetric.

   (b) $I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(1/2, 1/4, 1/4) = H(Y) - 3/2$. To maximize mutual information, we must maximize $H(Y)$ (by making Y as uniform as possible). Let $P[X = 0] = p$. We have $H(Y) = H(p/2 + (1-p)/4, p/4 + (1-p)/2, 1/4) = H((1+p)/4, (2-p)/4, 1/4)$. To achieve maximum entropy, we make $P[Y = 0] = P[Y = 1]$ by setting $1+p = 2-p$, i.e., $p = 1/2$. We then find that $C = H(3/8, 3/8, 2/8) - 3/2 \approx 0.061$ bits per channel use, achieved using a uniform input distribution.

3. (a) Let μ_i denote the steady state probability of state i. In steady-state, the probability "mass-flow" must be balanced across any cut in the graph. In particular, we find that $\mu_0/2 = \mu_1/2 + \mu_2$ and $\mu_1/2 = \mu_2$. Solving these equations, we find $(\mu_0 : \mu_1 : \mu_2) = (4 : 2 : 1)$. Since the sum of the probabilities is unity, we have $\mu_0 = 4/7$, $\mu_1 = 2/7$ and $\mu_2 = 1/7$. The entropy rate of the process is then

$$H(X) = H(X_1|X_0) = \sum_{i=0}^{2} H(X_1|X_0 = i)\mu_i = 1\cdot(4/7)+1\cdot(2/7)+0\cdot(1/7) = 6/7 \text{ bits/symbol.}$$

   (b) We can encode the starting state X_0 using a Huffman code with, e.g., $C(0) = 0$, $C(1) = 10$ and $C(2) = 11$. This takes an average of $1 \cdot 4/7 + 2 \cdot 3/7 = 10/7$ bits. We then encode the sequence of state values conditionally, depending on the current state. To be specific, we might use the code specified by the following table:

X_{i-1}	X_i	$C(X_i)$
0	0	0
0	1	1
1	0	0
1	2	1
2	0	ϵ

Here ϵ denotes the empty string; whenever the $X_{i-1} = 2$, then $X_i = 0$ with probability one, and so nothing needs to be sent.

A slight difficulty arises if we are encoding a finite sequence that happens to end in state 2, as in that case we have no way to distinguish between the sequence $w2$ and $w20$, where w is an arbitrary prefix. To get around this problem, we can append a single additional bit that will allow us to circumvent this issue. For example, the additional bit could be zero if the total number of input symbols is an even number and one otherwise.

This code is clearly uniquely decodable. The expected number of bits needed to encode a sequence of length n is

$$L_n = \underbrace{10/7}_{\text{first symbol}} + \underbrace{(n-1)[(4/7)(1) + (2/7)(1) + (1/7)(0)]}_{\text{next } n-1 \text{ symbols}} + \underbrace{1}_{\text{extra bit}} = 6n/7 + 11/7.$$

Thus per symbol, the expected encoding length is

$$\frac{1}{n} L_n = 6/7 + 11/7n \to 6/7 \text{ as } n \to \infty.$$

Thus, as $n \to \infty$, we approach the entropy rate of the process.

   (c) $101200 \mapsto 10, 0, 1, 1, \epsilon, 0, 0 = 1001100$ where the last 0 indicates that the sequence contained an even number of symbols.

4. (a) The contribution of $X^{(2)}$ is additional Gaussian noise. From our analysis of the Gaussian channel, we know that a $(2^{nR}, n)$ codebook can be successfully decoded with high probability when $R < \frac{1}{2}\log_2(1+\mathsf{SNR})$.
In the first decoding we have $\mathsf{SNR} = P_1/(P_2+N)$, so $R_1 < \frac{1}{2}\log_2(1+P_1/(P_2+N))$. In the second decoding, we have $\mathsf{SNR} = P_2/N$, so $R_2 < \frac{1}{2}\log_2(1+P_2/N)$.
(b) We now require $R_2 < \frac{1}{2}\log_2(1+P_2/(P_2+N))$ and $R_1 < \frac{1}{2}\log_2(1+P_1/N)$.
(c) We have

$$\begin{aligned}
R_1 + R_2 &< \frac{1}{2}\log_2(1+P_1/(P_2+N)) + \frac{1}{2}\log_2(1+P_2/N) \\
&= \frac{1}{2}\log_2((P_1+P_2+N)/(P_2+N)) + \frac{1}{2}\log_2((P_2+N)/N) \\
&= \frac{1}{2}\log_2((P_1+P_2+N)/N) \\
&= \frac{1}{2}\log_2(1+P/N)
\end{aligned}$$

We see that all rate sums $R_1 + R_2$ up to the channel capacity $C = \frac{1}{2}\log_2(1+P/N)$ are achievable.

5. (a) Let (q_0, q_1, q_2, q_3) be an arbitrary probability vector such that $0\cdot q_0 + 1\cdot q_1 + 2\cdot q_2 + 1\cdot q_3 \le w$. Let $(p_0, p_1, p_2, p_3) = (ab^0, ab^{-1}, ab^{-2}, ab^{-1})$, where $a = (2-w)^2/4$ and $b = (2-w)/w \ge 1$; then $p_i = p(i)$ as given in the problem statement. We have

$$\begin{aligned}
0 &\le D(q\|p) = \sum_{i=0}^{3} q_i \log(q_i/p_i) \\
&= -H(q_0,q_1,q_2,q_3) - [q_0\log(ab^0) + q_1\log(ab^{-1}) + q_2\log(ab^{-2}) + q_3\log(ab^{-1})] \\
&= -H(q_0,q_1,q_2,q_3) - \log a + \log(b)(0\cdot q_0 + 1\cdot q_1 + 2\cdot q_2 + 1\cdot q_3) \\
&\le -H(q_0,q_1,q_2,q_3) - \log a + w\log b \\
&= -H(q_0,q_1,q_2,q_3) - \log(ab^{-w})
\end{aligned}$$

Now

$$\begin{aligned}
-\log(ab^{-w}) &= -\log\left((1-w/2)^2\right) + w\log\left((1-w/2)/(w/2)\right) \\
&= -2(1-w/2)\log(1-w/2) - 2(w/2)\log(w/2) \\
&= 2\mathcal{H}(w/2)
\end{aligned}$$

Thus, $H(q_0,q_1,q_2,q_3) \le 2\mathcal{H}(w/2)$, and this maximum entropy is achieved with equality if and only if $q_i = p_i$ for all i, in which case the distribution achieves average Lee weight w.

(b) i. If we "guess" at the value of the source output, we will be correct 1/4 of the time, we will guess a symbol at Lee distance 1 from the true value 1/2 of the time, and we will guess a symbol at Lee distance 2 from the true value 1/4 of the time. Thus

$$D_{\max} = \frac{1}{4}\cdot 0 + \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 = 1.$$

If $\hat{X} = X$, $R_{\max} = I(X;\hat{X}) = H(X) = 2$ bits/symbol.

ii. For $0 \le D \le D_{\max} = 1$, we have

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(X) - H(X-\hat{X}|\hat{X}) \\ &\ge H(X) - H(X-\hat{X}) \\ &= 2 - H(X-\hat{X}) \end{aligned}$$

where $X - \hat{X}$ is taken modulo 4. To minimize $I(X;\hat{X})$ we must maximize $H(X-\hat{X})$ subject to the constraint that $E[\rho(X-\hat{X})] \le D$. However, this is precisely the optimization problem solved in (a) with $w = D$. We obtain $H(X-\hat{X}) \le 2\mathcal{H}(D/2)$; hence

$$I(X;\hat{X}) \ge 2 - 2\mathcal{H}(D/2). \tag{1}$$

To see that this lower bound is achievable, we construct a "modulo 4" channel, as shown in Fig. 1. The channel input is $\hat{X}$, the channel output is X, and the "noise" N, independent of $\hat{X}$, has probability mass function with $P[N=0] = (2-D)^2/4$, $P[N=1] = P[N=3] = D(2-D)/4$, $P[N=2] = D^2/4$. If we choose a uniform distribution for $\hat{X}$, then channel output distribution will also be uniform, as desired. In this case, $I(X;\hat{X}) = H(X) - H(N) = 2 - 2\mathcal{H}(D/2)$, showing that the lower bound

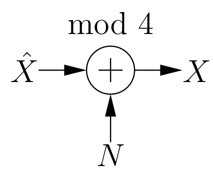


Figure 1: A modulo 4 channel.

(1) is achievable. Thus, for $0 \le D \le 1$,

$$R(D) = 2(1 - \mathcal{H}(D/2))$$

as plotted in Fig. 2.

iii. It is easy to see from Fig. 2 or by explicit calculation that $R(0) = 2(1-\mathcal{H}(0)) = 2 = R_{\max}$, and $R(1) = 2(1 - \mathcal{H}(1/2)) = 0$, as desired.

(c) Label the "Lee circle" with a Grey-code as shown in Fig. 3. Then the Lee distance between any two points is precisely the Hamming distance between their respective binary labels. This observation allows us to adapt a binary rate-distortion code for the quaternary source with Lee distortion: we simply encode the binary labels using the binary rate-distortion code. If R_b and D_b are the rate (bits per binary symbol) and Hamming distortion (per binary symbol), then, since each quaternary symbol is represented by two bits, we obtain $R = 2R_b$ (bits per quaternary symbol) and $D = D_b/2$ (per quaternary symbol). If the binary rate distortion code achieves the rate-distortion limit for the binary source with Hamming distortion, then $R_b = 1 - \mathcal{H}(D_b)$; in this case $R = 2(1 - \mathcal{H}(D/2))$, so this coding scheme also achieves the rate-distortion limit for the quaternary source with Lee distortion.

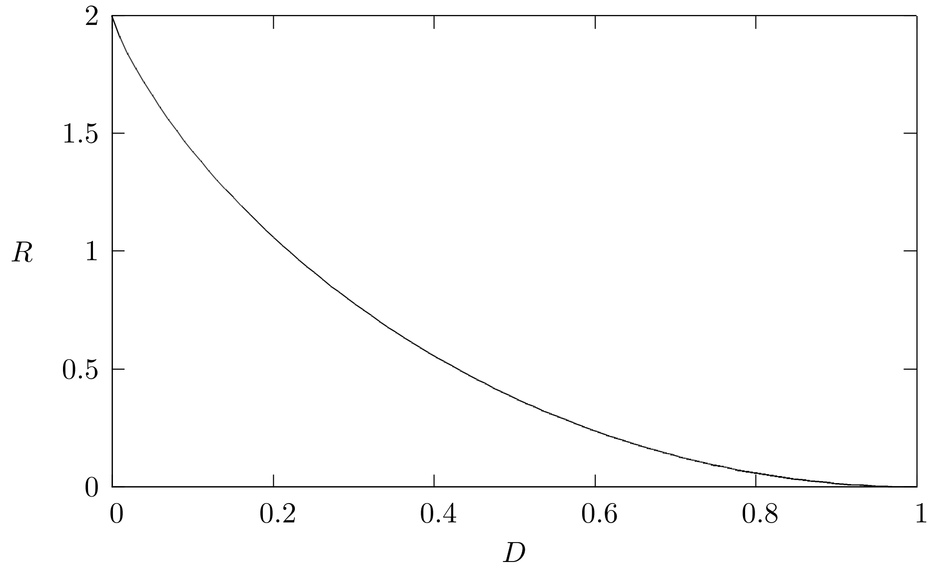


Figure 2: The rate-distortion function for a quaternary source with Lee distortion

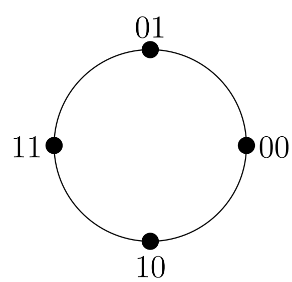


Figure 3: Hamming distance is the same as Lee distance with the binary labeling shown.