

ECE1502F — Information Theory
Final Examination Solutions
December 13, 2000

1. Short Snappers

- (a) $I(X; X) = H(X)$, the entropy of X .
- (b) False. Although $I(X; Y) = H(Y) - H(Y|X)$, maximizing $H(Y)$ does not necessarily maximize $I(X; Y)$. In fact, it may not be possible to achieve a uniform output distribution. An example is the binary erasure channel, where the capacity achieving input distribution yields an output distribution $((1 - \epsilon)/2, \epsilon, (1 - \epsilon)/2)$, which is not uniform unless $\epsilon = 1/3$.
- (c) This symmetric channel has capacity

$$\begin{aligned} C &= \log_2(3) - H(1/2, 1/4, 1/4) \\ &= \log_2(3) - (1/2) \log_2(2) - (1/4) \log_2(4) - (1/4) \log_2(4) \\ &= \log_2(3) - 3/2 \approx .085 \text{ bits/channel - use.} \end{aligned}$$

- (d) The channel transition matrix is

$$M = \begin{bmatrix} 1 - \epsilon & \epsilon & 0 \\ 0 & \epsilon & 1 - \epsilon \end{bmatrix}.$$

For this to define a weakly symmetric channel, the rows must be permutations of each other (they are), and the column sums must be equal, which occurs when $1 - \epsilon = 2\epsilon$, i.e., when $\epsilon = 1/3$.

- (e) We have

$$C = W \log_2 \left(1 + \frac{P}{N_0 W} \right) \text{ bits/s.}$$

If the SNR is 20 dB, we have $P/(N_0 W) = 100$, so $C \approx 20$ Kbps. If the SNR is 30 dB, we have $P/(N_0 W) = 1000$, so $C \approx 30$ Kbps. Thus, if the channel is as noisy as indicated, the telephone line capacity is on the order of 20 to 30 Kbps.

- (f) True. Let B be a code of length n and rate $R < C$ that achieves error probability ϵ . (By the coding theorem, we know that such a code exists for some length n (or longer) for any $\epsilon > 0$.) Form the code B' of length $n + 1$ by appending an extra symbol to each codeword of B , where the extra symbol is a one if the given codeword has an odd number of ones, and zero otherwise. Then every codeword of B' has an even number of ones.

Now, the error probability for B' is no greater than ϵ , since one decoding strategy for B' would be to use a decoder for B that simply ignores the additional symbol. The rate of B' is $nR/(n + 1)$. By choosing n large enough, we can make the rate of B' approach arbitrarily close to R , and hence arbitrarily close to C . Hence the given constraint does not affect the capacity-achieving ability of this family of codes.

- (g) False. Consider the binary symmetric channel with zero crossover probability, which has a capacity of one bit per channel use. The given family of codes cannot achieve this rate, since the given constraint limits the number of valid sequences.

Indeed, the set of valid sequences corresponds to the graph of Fig. 1. The adjacency matrix corresponding to this graph is

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

which has largest eigenvalue

$$\lambda = \frac{1}{3} \left(1 + (19 - 3\sqrt{33})^{1/3} + (19 + 3\sqrt{33})^{1/3} \right).$$

Thus the capacity of the constrained system is

$$\log_2 \lambda \approx 0.87915,$$

which is strictly less than one. Thus the given ensemble of codes cannot achieve the capacity of an arbitrary binary symmetric channel.

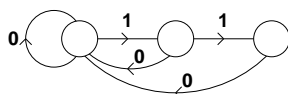


Figure 1: Constraint graph for binary sequences with forbidden subsequence 111.

2. A Phase-Noise Channel

- (a) We can write

$$\begin{aligned} I(X; Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X + Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \end{aligned}$$

Since $h(Z)$ is fixed, to maximize $I(X; Y)$, we should maximize $h(Y)$. The maximum possible $h(Y)$ is obtained when Y is distributed uniformly in $[-\pi, \pi)$. This can be achieved by choosing X to be distributed uniformly in $[-\pi, \pi)$.

To see this, observe that if $Y = X + Z$ then, in general,

$$\begin{aligned} f_Y(y) &= \int_x f_{X,Z}(x, y - x) dx \\ &= \int_x f_X(x) f_Z(y - x) dx, \end{aligned}$$

where the region of integration is $[-\pi, \pi)$, where $y - x$ is computed modulo 2π , and where the second equality follows from the fact that X and Z are independent. Setting $f_X(x) = 1/2\pi$, we obtain

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_x f_Z(y - x) dx \\ &= \frac{1}{2\pi} \int_z f_Z(z) dz \\ &= \frac{1}{2\pi}, \end{aligned}$$

where the second equality follows from the substitution $z = y - x$.

Thus Y is uniform in $[-\pi, \pi)$ when X is uniform in $[-\pi, \pi)$, and hence a uniform input distribution is capacity-achieving. The capacity of this phase noise channel is $C = \log_2(2\pi) - h(Z)$ bits per channel use.

(b) In this case

$$\begin{aligned} I(X; Y) &= h(Y) - h(Z) \\ &= \log_2(2\pi) - \log_2(2\pi/M) \\ &= \log_2(M) \text{ bits per channel use.} \end{aligned}$$

We can use as an input set any M -ary alphabet consisting of M points uniformly spaced by $2\pi/M$, for example: $\{-\pi, -\pi + 2\pi/M, -\pi + 4\pi/M, \dots, -\pi + (M - 1)\pi/M\}$. In this way we obtain a noiseless channel, since the sets of possible outputs corresponding to each possible input are disjoint.

Let us label the channel inputs as $0, 1, \dots, M - 1$, where label i maps to channel input $-\pi + 2i\pi/M$. Let us design a binary code.

When M is a power of 2, we can map the bits directly to the binary representation of i , and achieve a transmission rate of $\log_2 M$ bits per channel use directly, with zero probability of error.

When M is not a power of two, we can form a code of length n by taking a subset of the M^n possible vectors of channel inputs of length n . In general, we choose a subset of maximum possible size that is a power of two. In the limit as $n \rightarrow \infty$, we can achieve a transmission rate of $\log_2 M$ bits per channel use, with zero probability of error.

(c) We need only evaluate $h(Z)$. In nats, we have

$$\begin{aligned} h(Z) &= - \int_z f(z) \ln f(z) dz \\ &= - \int_z f(z) \ln \left(\frac{1}{2} \left(\frac{\alpha}{1 - e^{-\alpha\pi}} \right) e^{-\alpha|z|} \right) dz \\ &= \ln \left(\frac{2(1 - e^{-\alpha\pi})}{\alpha} \right) - \int_z f(z) \ln e^{-\alpha|z|} dz \\ &= \ln \left(\frac{2(1 - e^{-\alpha\pi})}{\alpha} \right) + \alpha \int_z f(z) |z| dz \\ &= \ln \left(\frac{2(1 - e^{-\alpha\pi})}{\alpha} \right) + 2\alpha \int_0^\pi f(z) z dz \quad (\text{since } f(z)|z| \text{ is an even function}) \end{aligned}$$

$$\begin{aligned}
&= \ln\left(\frac{2(1 - e^{-\alpha\pi})}{\alpha}\right) + \frac{\alpha^2}{1 - e^{-\alpha\pi}} \int_0^\pi z e^{-\alpha z} dz \\
&= \ln\left(\frac{2(1 - e^{-\alpha\pi})}{\alpha}\right) - \frac{\alpha^2}{1 - e^{-\alpha\pi}} \frac{1}{\alpha^2} (1 + \alpha z) e^{-\alpha z} \Big|_0^\pi \\
&= \ln\left(\frac{2(1 - e^{-\alpha\pi})}{\alpha}\right) + \frac{(1 - (1 + \alpha\pi)e^{-\alpha\pi})}{1 - e^{-\alpha\pi}}.
\end{aligned}$$

Since $h(Y) = \ln(2\pi)$, we find that

$$\begin{aligned}
C &= \ln(2\pi) - \ln\left(\frac{2(1 - e^{-\alpha\pi})}{\alpha}\right) - \frac{(1 - (1 + \alpha\pi)e^{-\alpha\pi})}{1 - e^{-\alpha\pi}} \\
&= \ln\left(\frac{\alpha\pi}{1 - e^{-\alpha\pi}}\right) - \frac{(1 - (1 + \alpha\pi)e^{-\alpha\pi})}{1 - e^{-\alpha\pi}} \text{ nats per channel use.}
\end{aligned}$$

3. Composite Channels

(a) The channel transition matrix for this channel is

$$M = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} = \begin{bmatrix} 1-2p(1-p) & 2p(1-p) \\ 2p(1-p) & 1-2p(1-p) \end{bmatrix}$$

This is a binary symmetric channel with crossover probability $2p(1-p)$, hence

$$C_{\text{cascade}} = 1 - \mathcal{H}(2p(1-p)).$$

(b) Clearly $C_{\text{recode}} \leq C_{\text{BSC}}(p) = 1 - \mathcal{H}(p)$, since the bit rate through the second channel is limited by this amount. Intuitively, since the recoder can decode/code reliably at all rates $R < C_{\text{BSC}}(p)$, we expect $C_{\text{recode}} = C_{\text{BSC}}(p)$. To prove all rates $R < C_{\text{BSC}}(p)$ are indeed achievable, fix a rate $R < C_{\text{BSC}}$ and $\epsilon > 0$. By definition, for n sufficiently large, there is some $(2^{nR}, n)$ code C^* for the BSC with maximal probability of error $\lambda^{(n)} \leq \epsilon/2$. We assume that the transmitter, recoder, and receiver all operate with the code C^* . For any message $w \in \{1, 2, \dots, 2^{nR}\}$, the probability of correct decoding at the receiver is at least as large as the probability that the recoder and the receiver both decode correctly, i.e.,

$$1 - \lambda_w \geq (1 - \epsilon/2) \cdot (1 - \epsilon/2) = 1 - \epsilon + \epsilon^2/4 \geq 1 - \epsilon.$$

Thus, $\lambda_w \leq \epsilon$. Since w was chosen arbitrarily, it follows that $\lambda^{(n)} \leq \epsilon$, and hence all rates $R < C_{\text{BSC}}$ are achievable.

(c) Clearly the overall capacity can be no greater than the smaller of the subchannel capacities, i.e.,

$$C \leq \min(C_{\text{BSC}}(p), C_{\text{BSC}}(q)).$$

Now, choose codes C_1^* and C_2^* for the two subchannels both of which operate at rate $R < \min(C_{\text{BSC}}(p), C_{\text{BSC}}(q))$. Then for n sufficiently large, the codes can be designed to operate at an arbitrarily small maximal error rate. Using the same argument as above, we see that the error rate at the receiver can be made arbitrarily small, and so all rates $R < \min(C_{\text{BSC}}(p), C_{\text{BSC}}(q))$ are achievable, and hence the capacity is $C = \min(C_{\text{BSC}}(p), C_{\text{BSC}}(q))$.

4. Entropy of Functions of Random Variables

- (a) Since $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$, and since $H(Y|X) = 0$ as Y is a deterministic function of X , it follows that

$$H(X) = H(Y) + H(X|Y).$$

Thus $H(X) = H(Y)$ if $H(X|Y) = 0$, which means that X is a deterministic function of Y , i.e., $f(x)$ is one-to-one, and hence invertible.

- (b) Let $\mathcal{S}_X = \{1, 2, \dots, m\}$, with $P[X = i] = p_i$ and let $Y = f(X)$, where

$$f(X) = \begin{cases} 0 & 1 \leq X \leq r \\ 1 & r+1 \leq X \leq m \end{cases}$$

Then $Y = 0$ with probability $A = \sum_{i=1}^r p_i$, and $Y = 1$ with probability $B = \sum_{i=r+1}^m p_i$. Furthermore, the conditional probability mass function for X , given that $Y = 0$, is

$$P[X = i|Y = 0] = \begin{cases} p_i/A & 1 \leq i \leq r \\ 0 & r+1 \leq i \leq m \end{cases}$$

and the conditional probability mass function for X , given that $Y = 1$, is

$$P[X = i|Y = 1] = \begin{cases} 0 & 1 \leq i \leq r \\ p_i/B & r+1 \leq i \leq m \end{cases}.$$

From the result of part (a), we have

$$\begin{aligned} H(X) &= H(Y) + H(X|Y) \\ &= H(Y) + P[Y = 0]H(X|Y = 0) + P[Y = 1]H(X|Y = 1), \end{aligned}$$

so that

$$\begin{aligned} H(p_1, p_2, \dots, p_m) &= H(A, B) + AH(p_1/A, p_2/A, \dots, p_r/A) \\ &\quad + BH(p_{r+1}/B, p_{r+2}/B, \dots, p_m/B) \end{aligned}$$

- (c) Let $Y = X_1 + X_2$, where X_1 and X_2 are discrete random variables. Note that Y is a deterministic function of X_1 and X_2 . Thus from (a) it follows that

$$H(X_1, X_2) = H(Y) + H(X_1, X_2|Y)$$

so

$$\begin{aligned} H(Y) &= H(X_1, X_2) - H(X_1, X_2|Y) \\ &\leq H(X_1, X_2) \quad (\text{equality iff } H(X_1, X_2|Y) = 0) \\ &\leq H(X_1) + H(X_2) \quad (\text{equality iff } X_1 \text{ and } X_2 \text{ are independent}) \end{aligned}$$

Thus $H(Y) = H(X_1) + H(X_2)$ if and only if X_1 and X_2 are independent, and X_1 and X_2 can be determined uniquely from their sum Y .

An example where this holds is obtained as follows. Take $X_1 \in \{0, 1\}$ and $X_2 \in \{0, 2\}$, independently and with uniform probability. Then $Y = X_1 + X_2$ is uniform in $\{0, 1, 2, 3\}$, and $H(Y) = 2 = H(X_1) + H(X_2)$. (Note that X_1 and X_2 are both uniquely determined from the value of Y .)

5. Channel Reduction

- (a) Let X denote the channel input, Y the channel output (before reduction) and Z the output of the reduced channel. Let y_1 and y_2 be output letters combined in the reduced channel, and let z be the corresponding output of the reduced channel. Assume that for all $x \in \mathcal{S}_X$ and some constant α ,

$$P[Y = y_1|X = x] = \alpha P[Y = y_2|X = x].$$

(This is equivalent to saying that the corresponding columns in the channel matrix are proportional.)

First, observe that

$$\begin{aligned} P[Y = y_1] &= \sum_{x \in \mathcal{S}_X} P[x]P[Y = y_1|X = x] \\ &= \alpha \sum_{x \in \mathcal{S}_X} P[x]P[Y = y_2|X = x] \\ &= \alpha P[Y = y_2]; \text{ therefore,} \\ P[Z = z] &= P[Y = y_1] + P[Y = y_2] \\ &= (1 + \alpha)P[Y = y_2]. \end{aligned}$$

Next, observe that

$$\begin{aligned} P[X = x|Y = y_1] &= P[Y = y_1|X = x]P[X = x]/P[Y = y_1] \\ &= \alpha P[Y = y_2|X = x]P[X = x]/(\alpha P[Y = y_2]) \\ &= P[Y = y_2|X = x]P[X = x]/P[Y = y_2] \\ &= P[X = x|Y = y_2]. \end{aligned}$$

Similarly, since

$$\begin{aligned} P[Z = z|X = x] &= P[Y = y_1|X = x] + P[Y = y_2|X = x] \\ &= (1 + \alpha)P[Y = y_2|X = x], \end{aligned}$$

we have

$$\begin{aligned} P[X = x|Z = z] &= P[Z = z|X = x]P[X = x]/P[Z = z] \\ &= (1 + \alpha)P[Y = y_2|X = x]P[X = x]/((1 + \alpha)P[Y = y_2]) \\ &= P[X = x|Y = y_2]. \end{aligned}$$

From this it follows that

$$H[X|Y = y_1] = H[X|Y = y_2] = H[X|Z = z].$$

Now, we can see that

$$\begin{aligned} H(X|Y) &= \sum_{y \in \mathcal{S}_Y} H(X|Y = y)P[Y = y] \\ &= \sum_{y \in \mathcal{S}_Y \setminus \{y_1, y_2\}} H(X|Y = y)P[Y = y] \end{aligned}$$

$$\begin{aligned}
& +H(X|Y = y_1)P[Y = y_1] + H(X|Y = y_2)P[Y = y_2] \\
= & \sum_{y \in \mathcal{S}_Y \setminus \{y_1, y_2\}} H(X|Y = y)P[Y = y] \\
& +H(X|Y = y_2)(P[Y = y_1] + P[Y = y_2]) \\
= & \sum_{z \in \mathcal{S}_Z} H(X|Z = z)P[Z = z] \\
= & H(X|Z).
\end{aligned}$$

From this it follows that

$$I(X; Y) = H(X) - H(X|Y) = H(X) - H(X|Z) = I(X; Z),$$

so mutual information, and hence capacity, is not reduced.

- (b) The transition matrix for the reduced channel is obtained from the original transition matrix by replacing the two columns corresponding to the symbols being combined with their sum. Thus if

$$M = \begin{bmatrix} 1/4 & 1/8 & 1/8 & 1/8 & 1/16 & 1/8 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 & 1/16 & 1/16 & 1/4 & 1/16 & 1/16 \\ 1/8 & 1/8 & 1/4 & 1/16 & 1/8 & 1/8 & 1/16 & 1/8 \end{bmatrix},$$

we combine columns 1, 4, and 7; columns 2 and 6; columns 3, 5, and 8 to obtain the reduced channel matrix

$$M' = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}.$$

The reduced channel is a symmetric channel with capacity

$$C = \log_2 3 - H(1/2, 1/4, 1/4) = \log_2 3 - 3/2 \text{ bits/channel-use.}$$