B8.4 Information Theory

Sheet 4 — MT23

Section A

1. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, X be a random variable on $\mathcal{X}$, and Y_1 and Y_2 be random variables on $\mathcal{Y}$. Conditioned on X, Y_1 and Y_2 are i.i.d..

   (a) Show that $I(X; Y_1, Y_2) = 2I(X; Y_1) - I(Y_1; Y_2)$.

   (b) Consider two DMCs of which (X, Y_1) and $(X, (Y_1, Y_2))$ are realisations. Prove that the capacity of the second DMC is at most twice that of the first.

   Solution:

   (a) We use the chain rule.

$$\begin{aligned} I(X; Y_1, Y_2) &= I(X; Y_2|Y_1) + I(X; Y_1) \\ &= I(X, Y_1; Y_2) - I(Y_1; Y_2) + I(X; Y_1) \\ &= I(Y_1; Y_2|X) + I(X; Y_2) - I(Y_1; Y_2) + I(X; Y_1) \\ &= 2I(X; Y_2) - I(Y_1; Y_2). \end{aligned}$$

   where we used $I(Y_1; Y_2|X) = 0$ since $Y_1 \perp Y_2$ conditioned on X, and $I(X; Y_2) = I(X; Y_1)$ since $p_{Y_2|X} = p_{Y_1|X}$.

   (b) By (a), we have $I(X; Y_1, Y_2) \leq 2I(X; Y_1)$, so

$$\sup_{p_X}\{I(X; Y_1, Y_2)\} \leq 2 \sup_{p_X}\{I(X; Y_1)\}.$$

2. Let $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and for each time $i \in \{1, \cdots, n\}$, we can use a DMC with transition matrix

$\mathcal{X} \backslash \mathcal{Y}$	0	1
0	$1 - q_i$	q_i
1	q_i	$1 - q_i$

   to transmit a symbol. This is an example of a time-varying discrete memoryless channel. Let $\mathbf{X} = (X_1, \cdots, X_n), \mathbf{Y} = (Y_1, \cdots, Y_n)$ with conditional pmf $\mathbb{P}(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x}) = \Pi_{i=1}^n \mathbb{P}(Y_i = y_i|X_i = x_i)$. Calculate $max_{p_{\mathbf{X}}} I(\mathbf{X}; \mathbf{Y})$.

Solution: As $I(\mathbf{X};\mathbf{Y}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X})$, we deal with the two terms separately.

$$\begin{aligned} H(\mathbf{Y}|\mathbf{X}=x) &= \sum_{i=1}^{n} H(Y_i|\mathbf{X}, Y_{i-1}, \cdots, Y_1) \\ &= \sum_{i=1}^{n} H(Y_i|\mathbf{X},) \\ &= \sum_{i=1}^{n} H(Y_i|X_i,) \\ &= \sum_{i=1}^{n} H(q_i); \end{aligned}$$

For the first term,

$$H(\mathbf{Y}) \leq \sum_{i=1}^{n} H(Y_i) \leq \sum_{i=1}^{n} \log(2) = n,$$

where the first equality hold iff Y_i are independent to each other, which is true when X_i are independent to each other; and the second equality hold iff Y_i is evenly distributed, which is true when X_i is evenly distributed.

So $\sup_{p_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n}(1 - H(q_i))$, and the capacity is obtained by the i.i.d. X_i with $\mathbb{P}(X_i = 0) = \mathbb{P}(X_i = 1) = 1/2$.

Section B

3. Consider the binary symmetric channel, i.e. $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and the transition matrix as below

$\mathcal{X}\backslash\mathcal{Y}$	0	1
0	$1-q$	q
1	q	$1-q$

Let $i \in \{1, \cdots, 16\}$ and consider a version of the Hamming $[7, 4, 3]_2$ code, as defined in lectures. That is, we take our 2^4 input codewords, enumerate them using binary digits $s = s_1s_2s_3s_4 \in \mathbb{F}_2^4$, and transmit

$$sG = s \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \in \mathbb{F}_2^7$$

Examples: $c(2) = 0001011$ since $s_1s_2s_3s_3 = 0001, c(5) = 0100110$ since $s_1s_2s_3s_4 = 0100$.

(a) Visualise this by drawing three intersecting circles. Put the first four bits into the regions intersecting at least two of these circles, and the parity bits in the remaining regions. Arrange the positions such that the sum of the four bits within each circle is even. Use this to find a good decoder $d : \mathcal{Y}^7 \mapsto \{1, \cdots, 16\}$, which will flip the minimal number of bits to restore even parity within each circle.

(b) Decode the outputs 1100101, 1000001.

(c) Calculate the error probabilities of this channel code.

(d) Calculate the rate of this channel code.

4. A channel with alphabet $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3, 4\}$ has emission matrix of the form

$$\mathbb{P}(y|x) = \begin{cases} 1/2 & \text{if } y = x \pm 1 \bmod 5 \\ 0 & \text{otherwise.} \end{cases}.$$

(a) Compute the capacity of this channel in bits

(b) The zero-error capacity of a channel is the number of bits which can be transmitted with zero probability of error. Find a block code that shows that the zero-error capacity of this channel is at least $(\log_2 5)/2$. [Hint: consider a code consisting of 5 codewords with block-length 2]

5. (a) Suppose X is a semi-Markov chain, that is, there exists an integer $k > 0$ such that

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, ...X_1 = x_1) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, ...X_{n-k} = x_{n-k})$$

for all $n > k$ and all $x_1, ..., x_n \in \mathcal{X}$. Show that the augmented process $(X_n, X_{n-1}, ..., X_{n-k})$ is a Markov chain, and describe its transition matrix.

(b) Suppose $\{X_t\}_{t\geq 1}$ is iid, and is encoded as a sequence $Z_1 Z_2 Z_3 ...$ by applying a Huffman coding algorithm. Show that the resulting process $\{Z_t\}_{t\geq 0}$ is not generally a Markov chain, but is a Markov chain if all probabilities are powers of two.

(c) Suppose $\{X_t\}_{t\geq 1}$ is a stationary Markov chain, and is encoded as a sequence of Blocks $(Z_1 ... Z_k)(Z_{k+1} ... Z_{2k})...$ using a Block Arithmetic Coding algorithm. Show that the process $\{(Z_{tk+1}, ..., Z_{(t+1)k}\}_{t\geq 1}$ is a Markov chain.

6. Using the one-step transition model of English characters (as in Problem Sheet 3), consider an erasure channel where each character could be replaced by $*$ with equal probability $q = 0.1$. Use the Viterbi algorithm to estimate the most likely source messages for the following observations, and comment on your results:

(a) `T*E_R*IN_*S_W*T`

(b) `JAB*ERWOCKY`

(c) `Q**EN_ELIZABETH`

(d) `*UEEN_ELIZABETH`

(e) `T******`

Section C

7. Let X be a time-invariant Markov chain with 2 states, with transition probabilities $P_{ij} = \mathbb{P}(X_{n+1} = x_j | X_n = x_i)$. Consider a DMC with emission matrix $M_{ij} = \mathbb{P}(Y_n = y_i | X_n = x_j)$. Suppose X is stationary, so $\mathbb{P}(X_n = x_i) = \mu_i = x_i \mu$, with the convention that X and Y both take values in the basis vectors (with x a row vector, y a column vector, as in Section 5.3 of the notes).

   (a) Give an example of a chain X and a non-invertible function g such that $\{g(X_t)\}_{t \geq 0}$ is a Markov chain.

   (b) Give a formula for the joint probability $\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$

   (c) Hence or otherwise, write down a necessary and sufficient algebraic condition on M and P under which Y is also a Markov chain.

Solution:

(a) A simple example is $g(x) \equiv 1$.

(b) We first write the probability of $(Y_i, X_i)_{i \leq 3}$:

$$\begin{aligned}
&\mathbb{P}(X_1 = x_1, Y_1 = y_1, ...) \\
&= \mathbb{P}(X_1 = x_1)\mathbb{P}(Y_1 = y_1 | X_1 = x_1)\mathbb{P}(X_2 = x_2 | X_1 = x_1)\mathbb{P}(Y_2 = y_2 | X_2 = x_2) \\
&\quad \times \mathbb{P}(X_3 = x_3 | X_2 = x_2)\mathbb{P}(Y_3 = y_3 | X_3 = x_3) \\
&= (x_1\mu)(x_1 M y_1)(x_1 P x_2^\top)(x_2 M y_2)(x_2 P x_3^\top)(x_3 M y_3)
\end{aligned}$$

We can therefore compute

$$\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) = \sum_{x_1, x_2, x_3} (x_1\mu)(x_1 M y_1)(x_1 P x_2^\top)(x_2 M y_2)(x_2 P x_3^\top)(x_3 M y_3)$$

(c) In order for Y to be a Markov chain, we require Y_3 to be independent of Y_1 given Y_2 (this is sufficient, as our problem is time-homogenous). We can calculate this quantity:

$$\begin{aligned}
&\mathbb{P}(Y_3 = y_3 | Y_1 = y_1, Y_2 = y_2) \\
&= \frac{\mathbb{P}(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)}{\mathbb{P}(Y_1 = y_1, Y_2 = y_2)} \\
&= \frac{\sum_{x_1, x_2, x_3} (x_1\mu)(x_1 M y_1)(x_1 P x_2^\top)(x_2 M y_2)(x_2 P x_3^\top)(x_3 M y_3)}{\sum_{x_1, x_2} (x_1\mu)(x_1 M y_1)(x_1 P x_2^\top)(x_2 M y_2)}
\end{aligned}$$

In particular, the right hand side should simplify to a function independent of y_1.

8. Consider a channel with binary input and output alphabets, and emission matrix

$$M = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

Find the capacity of the channel and the maximizing input probability distribution.

Solution: We write the information as a function of $p = \mathbb{P}(X = 1)$, that is

$$\begin{aligned} H(Y|X) = \mathbb{P}(X = 0) \times 0 + \mathbb{P}(X = 1) \times 1 = p \\ H(Y) = H(\mathbb{P}(Y = 1)) = H(p/2) \\ I(X;Y) = H(Y) - H(Y|X) = H(p/2) - p. \end{aligned}$$

Since $I(X;Y) = 0$ when $p = 0$ or $p = 1$, the maximum mutual information is obtained for some intermediate value of p. Using calculus, we see that

$$\frac{d}{dp} I(X;Y) = \frac{1}{2} \log \frac{1 - p/2}{p/2} - 1$$

which is zero when $p = 2/5$. So the capacity of the channel in bits is $H(1/5) - 2/5 \approx 0.722 - 0.4 = 0.322$.