Homework 6

1 7.2 Additive noise channel

2 7.4 Channel capacity

2.1 (a)

2.2 (b)

3 7.5 Using two channel at once

4 7.8 Z-channel

5 7.10 Zero error capacity

5.1 (a)

5.2 (b)

6 7.13 Erasures and errors in a binary channel

6.1 (a)

6.2 (b)

6.3 (c)

7 7.14 Channels with dependence between the letters

7.1 (a)

7.2 (b)

7.3 (c)

8 7.16 Encoder and decoder as part of the channel

8.1 (a)

8.2 (b)

8.3 (c)

8.4 (d)

9 7.23 Binary multiplier channel

9.1 (a)

9.2 (b)

10 7.28 Choice of channels

10.1 (a)

10.2 (b)

10.3 (c)

1 7.2 Additive noise channel

Denote $P(X=0)=p$

	Y=0	Y=1	Y=a	Y=a+1
X=0	$\frac{p}{2}$	0	$\frac{p}{2}$	0
X=1	0	$\frac{1-p}{2}$	0	$\frac{1-p}{2}$

Notice that the answer varies depending on the value of a.

Case 1: $a=0$

Then we get $X=Y$, which means $H(X|Y)=0$, then the capacity is $\max I(X;Y)=\max(H(X)-H(X|Y))=1bits$

Case 2: a = 1

	Y=0	Y=1	Y=2
X=0	$\frac{p}{2}$	$\frac{p}{2}$	0
X=1	0	$\frac{1-p}{2}$	$\frac{1-p}{2}$

$$\begin{aligned} H(X|Y) &= -\sum_x \sum_y p(x|y)\log p(x|y) \\ &= 2bits \end{aligned}$$

Case 3: a = -1

	Y=0	Y=1	Y=-1
X=0	$\frac{p}{2}$	0	$\frac{p}{2}$
X=1	$\frac{1-p}{2}$	$\frac{1-p}{2}$	0

$$\begin{aligned} H(X|Y) &= -\sum_x \sum_y p(x|y)\log p(x|y) \\ &= 2bits \end{aligned}$$

Case 4: $a \neq 0, \pm 1$

	Y=0	Y=1	Y=a	Y=a+1
X=0	$\frac{p}{2}$	0	$\frac{p}{2}$	0
X=1	0	$\frac{1-p}{2}$	0	$\frac{1-p}{2}$

From the table we know that given a value of Y, we know the value of X. Thus $H(X|Y)=0$.

$\max I(X;Y)=\max(H(X)-H(X|Y))=1bits$

2 7.4 Channel capacity

2.1 (a)

$$Z = \begin{cases} 1 & \text{with probability } \frac{1}{3} \\ 2 & \text{with probability } \frac{1}{3} \\ 3 & \text{with probability } \frac{1}{3} \end{cases}$$

Since $Y = X + Z \bmod 11$, X and Z are independent, we have

$$H(Y|X) = H(Z|X) = H(Z) - I(Z;X) = H(Z) = \log 3$$

Capacity $= \max I(X;Y) = \max(H(Y) - H(Y|X)) = \log 11 - \log 3 = \log \frac{11}{3} bits$

2.2 (b)

From the concavity of entropy and Jensen's Inequality, we know that the maximum of $H(Y)$ is reached when Y is uniformly distributed with probability $\frac{1}{11}$.

And when X is also uniformly distributed with probability $\frac{1}{11}$, then
$P(Y = i) = \frac{1}{3}P(X = i - 1 \bmod 11) + \frac{1}{3}P(X = i - 2 \bmod 11) + \frac{1}{3}P(X = i - 3 \bmod 11) = \frac{1}{11}, i = 0, 1, \ldots, 10$

which means that the maximum entropy of Y is reached.

3 7.5 Using two channel at once

From the expression of the channel $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1 \mid x_1) \times p(y_2 \mid x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ we know the joint distribution is

$$p(x_1, x_2, y_1, y_2) = p(x_1, x_2) p(y_1|x_1) p(y_2|x_2)$$

Which means

$$p(y_1, y_2|x_1, x_2) = p(y_1|x_1) p(y_2|x_2)$$

By taking expectation at two sides, we have

$$H(Y_1, Y_2|X_1, X_2) = H(Y_1|X_1) + H(Y_2|X_2)$$

Thus,

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= H(Y_1, Y_2) - H(Y_1, Y_2|X_1, X_2) \\ &= H(Y_1, Y_2) - H(Y_1|X_1, X_2) - H(Y_2|X_1, X_2) \\ &= H(Y_1, Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &\leq H(Y_1) + H(Y_2) - H(Y_1|X_1) - H(Y_2|X_2) \\ &= I(X_1; Y_1) + I(X_2; Y_2) \\ &\leq C_1 + C_2 \end{aligned}$$

The equality is obtained when Y_1 and Y_2 are independent, which means X_1 and X_2 are independent. And the distribution $p^*(X_1), p^*(X_2)$ maximize $I(X_1; Y_1), I(X_2, Y_2)$ respectively.

4 7.8 Z-channel

Denote $P(X=0)=p, P(X=1)=1-p$

$$(p \quad 1-p)\begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1+p}{2} & \frac{1-p}{2} \end{pmatrix}$$

$$H(Y|X) = \sum p(x)H(Y|X=x) = [p \cdot 0 + (1-p) \cdot 1] = 1-p \text{ bits}$$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= -[\frac{1+p}{2}\log\frac{1+p}{2} + \frac{1-p}{2}\log\frac{1-p}{2}] + p - 1 \\ &= p - \frac{1}{2}[(1+p)\log\frac{1+p}{2} + (1-p)\log\frac{1-p}{2}] \\ &= f(p) \end{aligned}$$

$$f'(p) = 1 - \log\frac{1+p}{1-p}$$

Let $f'(p) \geq 0 \Rightarrow p \leq \frac{3}{5}$, which means the maximum of $f(p)$ is reached at $p = \frac{3}{5}$ with maximum $f(\frac{3}{5}) \approx 0.3219$ bits.

And the corresponding distribution is

$$X \sim \begin{pmatrix} 0 & 1 \\ \frac{3}{5} & \frac{2}{5} \end{pmatrix}$$

5 7.10 Zero error capacity

5.1 (a)

$H(Y|X) = \sum p(x)H(Y|X=x) = \sum p(x) \cdot 1 = 1\text{bits}$

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= \log 5 - 1 \\ &\approx 1.3219 \text{ bits} \end{aligned}$$

5.2 (b)

According to the tips in the question, we consider code of 2 length. In this case, to make zero-capacity of the channel greater than 1 bits, we should have at least 5 codewords. Since the number of code words is integer, we could estimate that the maximal zero-capacity of the channel is $\frac{1}{2}\log 5 \approx 1.161$ bits.

We construct five codewords like this: $11, 23, 30, 42, 04$. Since there are totally 25 2-ary tuple, and it can be found that when send $11, 23, 30, 42, 04$, the message received exactly consist of remaining 20 2-ary tuple. Actually, each codeword is corresponding to a message group consisting of four 2-ary tuple, and each message group is disjoint.

In this case, the zero-capacity of the channel is exactly $\frac{1}{2}\log 5 \approx 1.161$ bits

6 7.13 Erasures and errors in a binary channel

6.1 (a)

Denote $P(X=0)=p, P(X=1)=1-p$

	Y=0	Y=e	Y=1
$X=0$	$1-\alpha-\epsilon$	α	ϵ
$X=1$	ϵ	α	$1-\alpha-\epsilon$

$$\begin{aligned} H(Y|X) &= \sum p(x)H(Y|X=x) \\ &= -p[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)+\alpha\log\alpha+\epsilon\log\epsilon] \\ &- (1-p)[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)+\alpha\log\alpha+\epsilon\log\epsilon] \\ &= -[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)+\alpha\log\alpha+\epsilon\log\epsilon] \end{aligned}$$

$$H(Y) = -(1-\alpha)\log\frac{(1-\alpha)}{2} - \alpha\log(\alpha) = H(\alpha)+1-\alpha$$

$$\begin{aligned} C = \max I(X;Y) = H(Y)-H(Y|X) &= H(\alpha)+1-\alpha+[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)+\alpha\log\alpha+\epsilon\log\epsilon] \\ &= 1-\alpha+[(1-\alpha-\epsilon)\log(1-\alpha-\epsilon)+\epsilon\log\epsilon-(1-\alpha-\epsilon+\epsilon)\log(1-\alpha)] \\ &= (1-\alpha)\left(1-H\left(\frac{1-\alpha-\epsilon}{1-\alpha},\frac{\epsilon}{1-\alpha}\right)\right) \end{aligned}$$

6.2 (b)

Set $\alpha=0$, we get $C=1-H(\epsilon)$, which is the capacity of symmetry binary channel.

6.3 (c)

Set $\epsilon=0$, we get $C=1-\alpha$, which is the capacity of the binary erasure channnel.

7 7.14 Channels with dependence between the letters

7.1 (a)

Let denote the distribution of input as $P(\{X_1,X_2\}=\{ij\})=p_{ij}$

$$\begin{aligned} I(X_1,X_2;Y_1,Y_2) &= H(Y_1,Y_2)-H(Y_1,Y_2 \mid X_1,X_2) \\ &= H(Y_1,Y_2)-0 \\ &= H(p_{11},p_{00},p_{01},p_{10}) \end{aligned}$$

7.2 (b)

$$C=\max I(X_1,X_2;Y_1,Y_2)=\max H(p_{11},p_{00},p_{01},p_{10}) \le \log|\mathcal{X}| = 2 \text{ bits}$$

7.3 (c)

The maximizing input distribution is $(p_{11},p_{00},p_{01},p_{10})=(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$

$$I(X_1;Y_1)=H(Y_1)-H(Y_1|X_1)$$

Since the input distribution is uniform distribution, we have $H(Y_1)=1$ bits, $H(Y_1|X_1)=1$ bits, thus $I(X_1;Y_1)=0$.

8 7.16 Encoder and decoder as part of the channel

8.1 (a)

Denote input symbol as X, output symbol as Y. Then $P(Y=1|X=0)=P(Y=0|X=1)=0.1$.

The crossover probability of $\{X^3\}$,$\{Y^3\}$ is $0.1^3+3\cdot 0.1^2\cdot 0.9=0.028=2.8\%$

8.2 (b)

From 7.13 we know that for a BSC with error probability $=0.028$, it's capacity is $1-H(0.028)\approx 0.8157$ bits, which corresponds to 0.2719 bits per transmission of the original channel.

8.3 (c)

Original capacity is $1-H(0.1)\approx 0.531$ bits

8.4 (d)

Let W denote the message input, X^n denote the codewords, Y^n denote the codewords after transmitted by channel, $\hat{W}$ denote the message decoded. Then according to the data processing inequality, we have

$$I(W;\hat{W})\leq I\left(X^n;Y^n\right)$$

therefore

$$C_W=\frac{1}{n}\max_{p(w)} I(W;\hat{W})\leq \frac{1}{n}\max_{p(x^n)} I\left(X^n;Y^n\right)=C$$

which implies the proposition is true.

9 7.23 Binary multiplier channel

9.1 (a)

Denote $P(X=0)=1-p, P(X=1)=p,$ then

	$Y=0$	$Y=1$
P	$1-\alpha p$	αp

Notice the fact that If $X=0$ then $Y=0$, then we have $H(Y|X=0)=0$

$$\begin{aligned} I(X;Y) &= H(Y)-H(Y\mid X) \\ &= H(Y)-\sum p(x)H(Y|X=x) \\ &= H(Y)-P(X=1)H(Z) \\ &= H(\alpha p)-pH(\alpha) \end{aligned}$$

Denote $f(p)=H(\alpha p)-pH(\alpha)=-[\alpha p\log(\alpha p)+(1-\alpha p)\log(1-\alpha p)-\alpha p\log\alpha-p(1-\alpha)\log(1-\alpha)]$

let $f'(p)\geq 0$, we have $p\leq \frac{1}{\alpha+(1-\alpha)^{\frac{\alpha-1}{\alpha}}}=\frac{1}{\alpha[1+(\alpha^{-\alpha}\cdot(1-\alpha)^{\alpha-1})^{\frac{1}{\alpha}}]}=\frac{1}{\alpha\left(2^{\frac{H(\alpha)}{\alpha}}+1\right)}$, $p^*=\frac{1}{\alpha\left(2^{\frac{H(\alpha)}{\alpha}}+1\right)}$ gives the maximum distribution on X. And the capacity is:

$$C=\max I(X;Y)=f(p^*)=\log\left(2^{\frac{H(\alpha)}{\alpha}}+1\right)-\frac{H(\alpha)}{\alpha}$$

9.2 (b)

Given that X and Z are independent, we have $I(X;Z) = 0$

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) = H(Y|Z) = p(Z=1)H(X) = \alpha H(p)$$

By the concavity of entropy function, the capacity is

$$C = \max I(X;Y,Z) = \alpha H(\frac{1}{2}) = \alpha$$

10 7.28 Choice of channels

10.1 (a)

Consider the following communication scheme:

$$X = \begin{cases} X_1 & \text{Probability } \alpha \\ X_2 & \text{Probability } (1-\alpha) \end{cases}$$

Let

$$\theta(X) = \begin{cases} 1 & X = X_1 \\ 2 & X = X_2 \end{cases}$$

Then we have $H(\theta) = H(\alpha)$.

Since the output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, θ is also a function of Y, we get a Markov chain: $X \to Y \to \theta$. ,and we have $I(X;\theta \mid Y) = 0$. Thus:

$$\begin{aligned} I(X;Y,\theta) &= I(X;\theta) + I(X;Y \mid \theta) \\ &= I(X;Y) \end{aligned}$$

$$\begin{aligned} I(X;Y) &= I(X;\theta) + I(X;Y \mid \theta) \\ &= H(\theta) - H(\theta \mid X) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2) \\ &= H(\alpha) + \alpha I(X_1;Y_1) + (1-\alpha) I(X_2;Y_2) \end{aligned}$$

Thus, it follows that

$$C = \sup_{\alpha} \{H(\alpha) + \alpha C_1 + (1-\alpha) C_2\}$$

Let $f(\alpha) = H(\alpha) + \alpha C_1 + (1-\alpha)C_2$, make $f'(\alpha) = 0$, we get $\alpha^* = \frac{2^{C_1}}{2^{C_1} + 2^{C_2}}$, and $2^C = 2^{f(\alpha^*)} = 2^{C_1} + 2^{C_2}$

10.2 (b)

From the intuitive explain of Capacity in the class lectures, we know that we could conceive C as the maximal effective number of bits to represent noise-free symbols, which further form disjoint set. Then if we have C_1 and C_2 effective bits respectively for channel 1 and 2, when we use them one by one, it's obviously the new effective number of noise-free symbols is $2^C = 2^{C_1} + 2^{C_2}$.

10.3 (c)

C_1 is the capacity of a BSC, from 7.13, we know its value is $1 - H(p)$

C_2 is obviously 0.

Then $C = \log(2^{1-H(p)} + 1)$