Solutions to Exercise Sheet 5

Exercise 1. Channel capacity: $C = \max_{p(x)} I(X:Y)$. There are three ways to calculate $I(X:Y)$:

1. $I(X:Y) = \sum_{x,y} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$.
2. $I(X:Y) = H(X) - H(X|Y)$.
3. $I(X:Y) = H(Y) - H(Y|X)$.

Note that for the (memoryless) additive noise channel where the input X and the noise Z are uncorrelated we can use the relation $H(Y|X) = H(X+Z|X) = H(Z)$. Therefore, for the calculation of the capacity we can use the equation

$$C = \max_{p(x)}\{H(Y)\} - H(Z), \tag{1}$$

where the second term $H(Z)$ does not depend on X (and by extension, on $p(x)$).

To compute the capacity as a function of a we need to consider three cases:

(I) $a = 0$: No noise is added, thus $Y = X$ and $H(Z) = 0$. The capacity is therefore $C = \max_{p(x)} H(p, 1-p) = 1$ bit ($P(X=0) = p, P(X=1) = 1-p$).

(II) $a > 1$: The output alphabet is given by $\mathcal{Y} = \{0, 1, a, 1+a\}$. For the input variable X we define the general probability distribution $P(X=0) = p, P(X=1) = 1-p$. Then we can compute probability distribution of the output Y:

$$\begin{aligned}
P(Y=0) = P(X=0)\cdot P(Z=0) &= \frac{p}{2},\\
P(Y=1) = P(X=1)\cdot P(Z=0) &= \frac{1-p}{2},\\
P(Y=a) = P(X=0)\cdot P(Z=a) &= \frac{p}{2},\\
P(Y=a+1) = P(X=1)\cdot P(Z=a) &= \frac{1-p}{2}.
\end{aligned}$$

We conclude that each output can be associated to a unique combination of input X and noise Z and thus, there is no error. We can recover the result in (I) by substituting the probability distribution $p(y)$ in equation (1):

$$C = \max_p\{-p\log_2(\frac{p}{2}) - (1-p)\log_2(\frac{1-p}{2})\} - 1 \text{ bit}.$$

This is maximized by $p = \frac{1}{2}$ for which $C = 1$ bit.

(III) $a = 1$: In this case, $\mathcal{Y} = \{0, 1, 2\}$. Similar to the reasoning above, if one obtains $Y = 0, 2$ then there is no error when guessing the X sent. However, $Y = 1$ corresponds to either $X = 0, Z = 1$ or $X = 1, Z = 0$. We find the output probability distribution $p(y) = \{\frac{p}{2}, \frac{1}{2}, \frac{1-p}{2}\}$. Substituting this in the capacity formula (1) we find that $C = \frac{1}{2}$ bits for $p = \frac{1}{2}$.

Exercise 2. This exercise can be solved in the same way as in Ex. 1: because the input X and noise Z are independent we can use Eq. (1) (warning: in general equation (1) is not valid!).

(a) We again parametrize the input probability distribution, but now, as the input takes 4 values we set it to $p(x) = \{a, b, c, d\}$ where $a + b + c + d = 1$ (alternatively one can include the constraint into the parametrization, so write $p(x) = \{a, b, c, 1 - a - b - c\}$). Note that $-1 \mod 4 = 3$, so the output alphabet reads $\mathcal{Y} = \{0, 1, 2, 3\}$. Now, we have to find the parameters a, b, c, d that maximize the output entropy $H(Y)$. We can express (similar to ex. **5-1**) $p(y)$ as a function of a, b, c, d

$$p(y) = \{\frac{a}{4} + \frac{c}{4} + \frac{d}{2}, \frac{b}{4} + \frac{c}{2} + \frac{d}{4}, \frac{a}{4} + \frac{b}{2} + \frac{c}{4}, \frac{a}{2} + \frac{b}{4} + \frac{d}{4}\}.$$

Now we try to find a, b, c, d such that the (optimal) uniform distribution $p(y) = \{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ is reached (warning: in general it may not be possible to achieve this! Usually we need to use Lagrange multipliers). We have 4 equations + 1 equation for the constraint $a + b + c + d = 1$ to solve (one equation is linearly dependent on the others). We find that $a = c$ and $b = d$. Namely, there is an infinite number of solutions and among them $a = b = c = d = 1/4$, *i.e.*, the uniform distribution for X is a solution.

(b) When we substitute the solution of (a) in Eq. (1) we find $C = \frac{1}{2}$ bits.

(c) We need to sum both noises: $Z_{total} = Z_1 + Z_2$ and again follow a calculation like in (a). One obtains the new probability distribution by noting that $(-2 \mod 4) = (2 \mod 4) = 2$. Z_{total} takes the values: $\{-1, 0, 1, 2\}$ with probabilities $\{\frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{8}\}$. We have thus $H(Z) = H(\frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{8})$ and we find $C = \log_2 4 - 1.91 = 0.09$ bits.

(d) $C_{total} = C_1 + C_2 = 1$ bit. (The capacity is additive!)

Exercise 3.

(a) If $p(X = 1) = p$ and $p(X = 0) = 1 - p$, we obtain:

$$I(X : Y) = H(1 - \frac{p}{2}, \frac{p}{2}) - p$$

Taking into account that:

$$\frac{\partial H(x, 1 - x)}{\partial x} = \log_2 \frac{1 - x}{x}$$

The maximum is found for $p^* = \frac{2}{5}$ and $C = I_{p=\frac{2}{5}}(X, Y) = \log_2 5 - 2$ bits.

(b) The binary symmetric channel has capacity $C = 1 - H(\alpha)$ bits, where α is the error rate of the channel, because $I(X : Y) = H(Y) - \sum p(x)H(Y|X = x) = H(Y) - H(\alpha) \leq 1 - H(\alpha)$ bits. This capacity is equal to the one derived above for $\alpha^* \approx 0.179$ and $1 - \alpha^* \approx 0.821$.

(c) We have $C = \max_{p(x)} I(X : Y)$, where $I(X : Y) = H(Y) - H(Y|X)$. The entropy at the output is given by $H(Y) = H(1 - p \cdot q, p \cdot q)$ and the conditional entropy reads $H(Y|X) = pH(q, 1 - q)$. $I(X : Y)$ is maximal if $\frac{\partial I(X:Y)}{\partial p} = 0$, which implies

$$q \cdot \log_2 \frac{1 - p \cdot q}{p \cdot q} = H(q, 1 - q).$$

To simplify notations we write $H(q, 1 - q) = H$. The distribution p that maximizes $I(X : Y)$ is

$$p = \frac{1}{q(1 + 2^{(\frac{H}{q})})}.$$

We finally obtain the capacity of the channel:

$$C = \log_2(1 + 2^{(\frac{H}{q})}) - \frac{H}{q}. \tag{2}$$

To check consistency we can test Eq. 2 for $q = 0.5$. Since $H(q = 0.5) = 1$ we confirm the result of (a), *i.e.* $C(q = 0.5) = \log_2 5 - 2$ bits.

Exercise 4.

1. It does not attain the capacity because the uniform distribution does not maximize the mutual information of the channel of Ex. 3.

2. For a probability distribution $p(x)$, the maximal transmission rate R is bounded from above by $I(X:Y)$. For the channel of exercise 5-3: $R_{p=1/2} < I_{p=1/2}(X:Y) = 0.3113$ bits.

Exercise 5.

1. $C = \max_{p(x)} I(X:Y)$ and $\tilde{C} = \max_{p(x)} I(X:\tilde{Y})$.
   We have $I(X:Y,\tilde{Y}) = H(X:Y) + H(X:\tilde{Y}|Y)$,
   and $I(X:Y,\tilde{Y}) = H(X:\tilde{Y}) + H(X:Y|\tilde{Y})$.
   Since $H(X:Y|\tilde{Y}) \geq 0$ and $H(X:\tilde{Y}|Y) = 0$ (see Exercise 3 in Sheet 2), we deduce that $I(X:\tilde{Y}) \leq I(X:Y)$. Thus, $\tilde{C} > C$ is impossible.

2. The requirement implies $H(X:Y|\tilde{Y}) = 0$. The channel satisfying this is given by $X \to \tilde{Y} \to Y$. This is only possible if $Y \longleftrightarrow \tilde{Y}$, i.e. iff $\tilde{Y} = f(Y)$ is a bijective function.