Quantum Information Theory
Solutions 2

Exercise 2.1 Information and Description Length

a) *How much information is gained upon learning (i) the state of a flipped coin; (ii) the states of two flipped coins; (iii) the outcome of the roll of a four-sided die?*

Assuming the coin is fair, we learn one bit from one flipped coin, two bits from two coins, and also two bits from the (fair) four-sided die. If the states are equally-likely, the entropy is the log of the number of them.

b) *How much information is gained when the odd ball and its weight are identified?*

There are 24 possible states as any of the 12 balls could be either light or heavy. Hence we learn $\log 24$ bits.

c) *How much information is gained on the first step if six balls are weighed against the other six? How much is gained by first weighing four against another four, leaving the rest aside?*

Weighing six against six rules out half of the possibilities. Suppose we weigh 1-6 versus 7-12. If the 1-6 pan comes up lighter, then there are only 12 possibilities remaining: either one of 1-6 is light or one of 7-12 is heavy. The opposite holds for the opposite result, so we've shrunk the list of possible states from 24 to 12, an information gain of one bit.

In contrast, weighing two arbitrary sets of four balls gives more information. Suppose we weigh 1-4 versus 5-8. If they balance, then one of 9-12 is the odd ball and is either light or heavy, for a total of 8 possibilities. If 1-4 is lighter, then either one of 1-4 is light or one of 5-8 is heavy, again 8 possibilities. The opposite states are possible if 1-4 is heavy. Therefore, no matter the outcome the list of possibilities has shrunk from 24 elements to 8, a gain of $\log 3$ bits.

d) *For your prospective weighing strategy, draw a tree showing the possible outcomes of the chosen weighing and what weighing is to be performed next. At each node, how much information has been gained and how much remains to be gained?*

The goal in designing a strategy is to maximize the information gain at every step. Since there are 24 possible states and each weighing could in principle have three equally-likely outcomes, it will take at least 3 weighings to determine which ball is odd and its weight ($3^2 < 24 < 3^3$).

Here's a method showing that it can indeed be accomplished in three weighings. First, pick two sets of four at random and compare them. If they are of equal weight, the odd ball is in the remaining four. Now pick three of these and weigh them against three balls known to be normal. If both sets are of equal weight, the oddball is the remaining one, and weighing it against a normal ball will determine if its heavy or light. If the sets are not equal, we have learned whether the oddball is heavy or light and it can only be one of the remaining three. Weighing one of them against the other determines which is which. Going back to the case in which the two sets of four are unequal, we know that the oddball is either light and one of the four light-weighing balls or heavy and one of the heavy-weighing set. Now break the light set into two sets of two and then add a random possibly heavy ball to each group. Weigh them. If they are of equal weight, the oddball is one of the remaining two, and since we know whether each one could be light or heavy, one further measurement sufficies to determine the oddball. If the light set turns out to be the two possibly light balls and one possibly heavy, then we know that it could not have been the possibly heavy ball, nor could it have been either of the two possibly light balls in the other set. But it could have been the two possibly light balls in the light set, or the possibly heavy ball in the heavy set. Now there are three possibilities remaining, and we only need to measure the two possibly light balls to determine which is which.

Exercise 2.2 Mutual Information

a) *Compute the mutual information between your guess and the actual weather, and do the same for your grandfather. Remember that all your grandfather knows is that it rains on* 80% *of the days. You know that as well and you also listen to the weather forecast and know that it is right* 80% *of the time and is always correct when it predicts rain.*

The mutual information is given by $I(X:Y) = H(X) - H(X|Y)$. Let us call your grandfather G, you Y and the actual weather W. We may also assume you followed the radio forecast (which we saw was an optimal strategy) and we will hold on to the notation $\hat{R}, \hat{S}$ for guesses (both yours and your grandfather's). Then we have, for the grandfather

$$\begin{aligned}
H(W) &= -P(R)\log P(R) - P(S)\log P(S)\\
&= -0.8\log 0.8 - 0.2\log 0.2\\
H(G) &= -P(\hat{R})\log P(\hat{R}) - P(\hat{S})\log P(\hat{S})\\
&= -1.\log 1. - 0. = 0\\
H(GW) &= -P(\hat{R}R)\log P(\hat{R}R) - P(\hat{S}R)\log P(\hat{S}R) - P(\hat{R}S)\log P(\hat{R}S) - P(\hat{S}S)\log P(\hat{S}S)\\
&= -0.8\log 0.8 - 0 - 0 - 0.2\log 0.2\\
H(W|G) &= H(GW) - H(G)\\
&= -0.8\log 0.8 - 0.2\log 0.2\\
I(W:G) &= H(W) - H(H|G)\\
&= 0.
\end{aligned}$$

For your case we will calculate the conditional entropy directly,

$$\begin{aligned}
H(W) &= -0.8\log 0.8 - 0.2\log 0.2\\
H(W|Y) &= -P(\hat{R})\left[P(R|\hat{R})\log P(R|\hat{R}) + P(S|\hat{R})\log P(S|\hat{R})\right]\\
&\quad - P(\hat{S})\left[P(R|\hat{S})\log P(R|\hat{S}) + P(S|\hat{S})\log P(S|\hat{S})\right]\\
&= -0.6[1\log 1 + 0] - 0.4[0.5\log 0.5 + 0.5\log 0.5]\\
&= -0.4\log 0.5\\
I(W:Y) &= H(W) - H(H|Y)\\
&= -0.8\log 0.8 - 0.2\log 0.2 + 0.4\log 0.5\\
&= 0.32.
\end{aligned}$$

b) *What would your strategy be? And your grandfather's? After N days, what is the expected gain for each of you? What is the probability that he finishes with more money than you?*

Here your optimal strategy and your grandfather's optimal strategy are the strategies that maximize the total amount of money you will each have after N days.

Let us first compute your grandfather's expected gain after N days. If your grandfather bets all his money on rain every evening (since this event is most likely for him each day), his expected amount of money is:

$$\langle \pounds_G \rangle = 0.8^N 2^N = 1.6^N.$$

As for you, how much you bet when the radio predicts a sunny day has no impact on your expected gain (because you will guess correctly 50% of the time: $P_{R|\hat{S}} = P_{S|\hat{S}} = 50\%$). So you should choose to keep your money when the forecast is sun (so that you don't lose with probability 50%). Your expected amount of money is

$$\begin{aligned} \langle \pounds_Y \rangle &= \sum_{k=0}^{N} P_{\hat{R}}(k)\ 2^k \\ &= \sum_{k=0}^{N} \binom{N}{k} 0.6^k\ 0.4^{N-k}\ 2^k \\ &= 1.6^N, \end{aligned}$$

where $P_{\hat{R}}(k)$ is the probability that the radio predicts rain on exactly k of the N days.

The expected amount of money for both of you is the same!

If there is rain every day then $k = N$, and your grandfather will have 2^N pounds. What is the probability that you have less money than your grandfather in this case? You will have to have a prediction of rain for each day. This is $P_{\hat{R}|R}^N = 0.75^N$. Therefore the probability that your grandfather makes more money than you when $k = N$ is $1 - (0.75)^N$.

If there is one or more days of sun ($k \neq N$), then your grandfather will have lost all his money. You will never lose all your money (either you don't bet when the forecast is sunny, or you are 100% sure that when you bet all your money on rain when the forecast is rain).

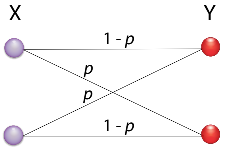
Therefore, your grandfather has more money than you with probability $1-0.8^N(1-0.75^N)$.

In other words, if you play this game for many days, you're extremely likely to have more money than your grandfather.

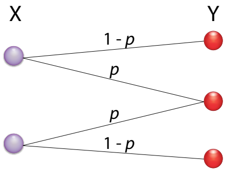
If you chose to bet randomly when the radio predicts sun, then the above probability of having more than your grandfather will decrease, because you could lose all your money with this strategy.

Exercise 2.3 Channel capacity

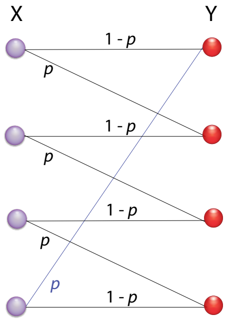
a) *The asymptotic channel capacity is given by $C = \max_{P_X} I(X:Y)$. Calculate the asymptotic capacities of the first two channels depicted above.*



(a) Binary Symmetric Channel



(b) Symmetric Erasure Channel



(c) Yet Another Channel

The capacity of the binary symmetric channel evaluates to

$$\begin{aligned}
C &= \max_{P_X} I(X:Y) = \max_{P_X} H(Y) - H(Y|X) \\
&\qquad\qquad\quad = \max_{P_X} H(Y) + \sum_{x,y} P_X(x) P_{Y|X=x}(y) \log P_{Y|X=x}(y) \\
&\qquad\qquad\quad = \max_{P_X} H(Y) - \sum_{x} P_X(x) \; H_{\text{bin}}(p) \\
&\qquad\qquad\quad = \max_{P_X} H(Y) - \; H_{\text{bin}}(p) \\
&\qquad\qquad\quad = 1 - H_{\text{bin}}(p),
\end{aligned}$$

where $H_{\text{bin}}(p)$ is the *binary entropy*, i.e. the entropy of the probability distribution $(p, 1-p)$,

$$H_{\text{bin}}(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

To maximise $H(Y)$ we chose the uniform distribution on the input, $P_X^u = (\frac{1}{2}, \frac{1}{2})$ (see part b).

Similarly, for the symmetric erasure channel, we have

$$\begin{aligned}
C &= \max_{P_X} H(Y) - H(Y|X) \\
&= \max_{P_X} H(Y) - H_{\text{bin}}(p) \\
&= 2\frac{1-p}{2} \log \frac{2}{1-p} + p \log \frac{1}{p} - H_{\text{bin}}(p) \\
&= 1 - p.
\end{aligned}$$

b) *Consider N possible probability distributions as input to a general channel, $\{P_X^i\}_i$, with the property that $I(X:Y)_{P^i} = I(X:Y)_{P^j}, \forall i, j$. Suppose you choose which distribution to use for the input by checking a random variable, B, with possible values $b = \{1, \ldots, N\}$. Show that $I(X:Y|B) \leq I(X:Y)$.*

We have

$$\begin{aligned}
I(X:Y|B) &= H(Y|B) - H(Y|XB) \\
&= H(Y|B) - H(Y|X) \qquad (*) \\
&\leq H(Y) - H(Y|X) \qquad (**) \\
&= I(X:Y),
\end{aligned}$$

where $^{(*)}$ stands because B is just a label that tells us which probability distribution P_X^i we used, so knowing X is as good as knowing X and B, in the sense that $H(Y|XB) = H(Y|X)$, and $^{(**)}$ comes from the data-processing inequality, $H(Y|B) \leq H(Y)$ (which in lay terms says that extra information cannot hurt).

How can you use that to find the probability distribution P_X that maximises the mutual information for symmetric channels? **Hint:** *consider $\left\{P_X^i\right\}_i$ permutations of P_X^1.*

For symmetric channels, the mutual information between input and output is invariant under permutation of the inputs (that's how they are defined). Look for instance at the symmetric erasure channel. The input distribution $P_X^1 = (0.75, 0.25)$ yields the same mutual information as $P_X^2 = (0.25, 0.75)$.

Not knowing which permutation of P_X was used in the input is equivalent to take a uniform mixture over all possible permutations of P_X. Conveniently, such mixture gives us the uniform distribution:

$$P_X = \begin{pmatrix} P_X(x_1) \\ P_X(x_2) \\ \vdots \\ P_X(x_N) \end{pmatrix}, \qquad \left\{P_X^i\right\}_{i=1,\dots,N!} \text{ permutations of } P_X,$$

$$\sum_i^{N!} \frac{1}{N!} P_X^i = \frac{1}{N!} \begin{pmatrix} (N-1)!\; P_X(x_1) + (N-1)!\; P_X(x_2) + \cdots + (N-1)!\; P_X(x_N) \\ (N-1)!\; P_X(x_1) + (N-1)!\; P_X(x_2) + \cdots + (N-1)!\; P_X(x_N) \\ \vdots \\ (N-1)!\; P_X(x_1) + (N-1)!\; P_X(x_2) + \cdots + (N-1)!\; P_X(x_N) \end{pmatrix}$$

$$= \frac{1}{N} \begin{pmatrix} \sum_i P_X(x_i) \\ \sum_i P_X(x_i) \\ \vdots \\ \sum_i P_X(x_i) \end{pmatrix} = \begin{pmatrix} 1/N \\ 1/N \\ \vdots \\ 1/N \end{pmatrix}.$$

This means that for any input distribution P_X the mutual information always increases if instead you use the uniform distribution. Here, $I(X:Y|B)$ is the mutual information knowing you used P_X (take $B=1$), and $I(X:Y)$ is the mutual information for the uniform distribution. Conclusion: for symmetric channels the mutual information is maximised if one takes the uniform distribution as input.

c) Using part *b*), we choose the uniform distribution on X and calculate the capacity:

$$\begin{aligned} C &= \max_{P_X} I(X:Y) = I(X:Y)_{P_X^u} = H(Y) - H(Y|X) \\ &= -\sum_y P_Y(y) \log P_Y(y) + \sum_{x,y} P_{XY}(x,y) \log P_{Y|X}(y) \\ &= -\sum_y \left(\sum_x P_X(x) P_{Y|X=x}(y) \right) \log \left(\sum_x P_X(x) P_{Y|X=x}(y) \right) + \sum_{x,y} P_X(x) P_{Y|X=x}(y) \log P_{Y|X=x}(y) \\ &= -4 \cdot \frac{1}{4} \log \left(\frac{1}{4} \right) + 4 \cdot \frac{1}{4} H_{\text{bin}}(p) = 2 - H_{\text{bin}}(p). \end{aligned}$$