ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quantum Information Theory
Tips 3

HS 12
Lea and Norm

Exercise 3.1 Channel capacity

Channels! A channel is a rather intuitive concept. Think of a noisy telephone line from the thirties. The question here is: how do we characterise the telephone line? We want to know how well a person on the other side will understand us when we phone. The relevant parameters cannot be the input sounds — those will change each time we use the channel. We are more interested in how reliably the telephone will reproduce each sound input: each time I say "aye", what is the probability that the sound that arrives the other side is "aye" and not "nay"? In other words, what is the probability of getting an "aye" *conditioned* on the fact that I input an "aye"? You can see where this is leading. A channel is fully characterised by the set of conditional probabilities of the outputs given each of the inputs. Pages 10–11 of the script have details and a much more precise formulation of what a channel is.

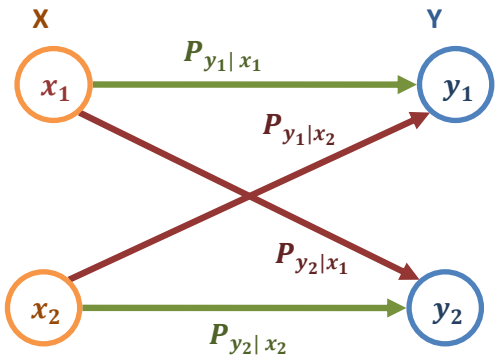

Figure 1: A channel with two inputs x_1 and x_2 and two outputs is defined by the conditional probabilities $P_{y_i|x_j}$.

You may see that in exercise 2.2 we had a channel — expect that in that case we also fixed the probabilities of each input. Now that you have characterised your telephone line with all the conditional probabilities, you want to find a way of quantifying how reliable it is. One way of doing this is to ask "I want to send a message through this channel with only a negligible probability of error. How long can that message be?" In the iid limit (i.e. you use the channel many times), the answer is the *capacity* of the channel. This is explained in detail in pages 18–22 of the script. Here as usual I will just try to give a feeling of its meaning.

You have seen that the mutual information gives us an amount of how correlated two things are. That is precisely what we want of a channel — the more correlated the input and output are, the better the channel. The quality (or capacity) of a channel should be related to the mutual information between input and output.

There is one free parameter in a channel, which is the probability distribution on the inputs. We can use it to maximise the certainty that our message will be well received by *encoding* our message. For instance, imagine a channel that transmits "ayes" correctly with 99% of probability but fails at transmitting "nays" 30% of the time. We may use redundancy to ensure our "nays" will be understood as such, by saying "nay nay nay" for each "nay" intended. The person in the other side will *decode* any sequence of two or three "'nays" (and one or none "aye") as a single "nay".

So, as we can use P_X to maximise the fidelity of the channel, the final capacity is given by

$$C = \max_{P_X} I(X:Y). \tag{1}$$

In part *a*) you have to apply this to two simple channels. You will find that the distribution P_X that maximises the mutual information is the uniform distribution. In part *b*) you are going to prove that that is the case for all symmetric channels.

You start by considering N probability distributions for the input, $P_X^1, \dots P_X^N$, such that $I(X:Y)_{P^i} = I(X:Y)_{P^j}, \forall i,j$. As an example you can think of a symmetric channels, where a permutation of the input probability distribution does not change the mutual information between input and output: $P_X^2, \dots P_X^N$ could be permutations of P_X^1.

Now suppose that Joanna chooses which probability distribution she will use for an input by picking a ball from a bag at random. Formally, this is expressed by a random variable B that can take values $b = 1, \dots N$ (assume a uniform probability distribution on the outcomes of B).
Now you compare the mutual information between input and output of the channel for Joanna, who knows which ball she picked—and therefore which P_X^i she chose as input, $I(X:Y|B)$, and someone who does not know which distribution she chose, $I(X:Y)$. Use properties of the conditional entropies to prove this; in particular, do not forget that knowing more cannot hurt ($H(A|B) \leq H(A)$), and that the conditional probabilities that define the channel are fixed.
You should get that $I(X:Y|B) \leq I(X:Y)$, i.e. one is always better if one does not know which P_x^i was used. "Not knowing which distribution was used" is the same as admitting that a uniform mixture of those distributions was used, i.e. P_X^1 with probability $1/N$, P_X^2 with probability $1/N$, etc. But what does that mean for symmetric channels? When all the P_x^i are permutations of each other, what is their uniform mixture? Up to you to work out!

Exercise 3.2 Smooth entropies in the i.i.d. limit

In this exercise you will see that in the i.i.d. limit the smooth min-entropy is equivalent to Shannon entropy. If you remember the exercises from last week, we said that Shannon entropy could be seen as an *average* uncertainty about an experiment (ie. particularly relevant in the i.i.d. limit), while the min entropy was a measure of the probability of making a correct guess about the outcome of an experiment and the max entropy gave us an upper bound for the memory size necessary to store all possible outcomes (the smooth versions allowed us to optimise these quantities if we tolerated a small error probability).
These three measures of uncertainty do not seem to be highly correlated – we saw some examples of how they could be very different for the same probability distributions – so it may be surprising that they converge when you repeat an experiment many times in the same conditions. The secret for that is that the law of large numbers modulates i.i.d. probability distributions in a way that favours similarities among the different entropy measures. Intuitive example: suppose the experiment "try a grape" can have the outcomes "sweet grape" with probability 90% and "bitter grape" with probability 10%. If you try a thousand grapes it will be very unlikely that they are all sweet or all bitter; you expect that around a hundred of them will be bitter and the rest will be sweet. In other words, you expect a *typical* pack of a thousand grapes to have around $1000 \times 90\% = 900$ good grapes and $1000 \times 90\% = 100$ bitter grapes.
In general, when an experiment is repeated many times in i.i.d. conditions it is very hard to find an atypical sequence, ie. one where the number of outcomes of each type is not proportional to the probability of that outcome in a single instance of the experiment (see Fig. 2).

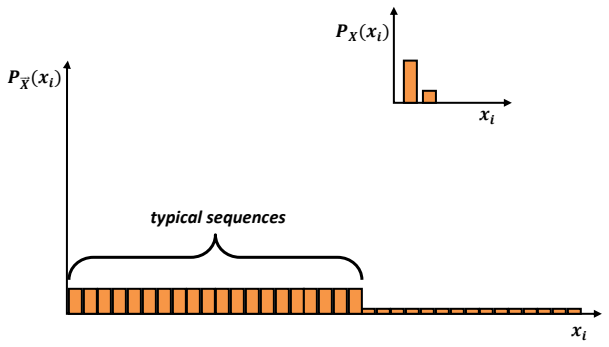

Figure 2: Almost all the weight of an i.i.d. probability distribution is spread among typical sequences. If the individual probability distribution of an experiment represented by X is $P_X = (P_X(x_1), P_X(x_2), \dots, P_X(x_k))$, then a typical sequence of n i.i.d. repetitions of the experiment has *approximately* $nP_X(x_1)$ outcomes of type x_1, $nP_X(x_2)$ outcomes of type x_2, etc.

This way a probability distribution of i.i.d. experiments for large numbers looks flat for typical sequences with a tail of very very unlikely atypical sequences. This tail can be ignored if we tolerate a small error probability (that we will see that can be narrowed down to zero in the case of infinite i.i.d. repetitions), and what is left is an uniform distribution

on typical sequences. We know that for uniform distributions min-, max- and Shannon entropies are the same. Of course you will have to say all of this in a more precise way and without grapes.
Here goes a suggestion to resolve this. Start by defining the set of typical sequences as that of the outputs $\vec{x} = (x_1, x_2, \ldots, x_n)$ whose average *surprise content* is close to the Shannon entropy of a single instance of the experiment, ie. of the random variable X,

$$\mathcal{G}_\nu = \left\{ \vec{x} \in \mathcal{X}^{\times n} : \left| \left[\frac{1}{n} \sum_i - \log P_X(x_i) \right] - H(X) \right| < \nu \right\}. \tag{2}$$

Use the weak law of large numbers that we have seen in exercise series 1 to show that as the number of repetitions of the experiment goes to infinity almost all outcomes belong to the typical set, *i.e.*

$$\lim_{n\to\infty} P_{\vec{X}}[\mathcal{G}_\nu] = \lim_{n\to\infty} P_{\vec{X}}[\vec{x} \in \mathcal{G}_\nu] = 1 \tag{3}$$

for any bound $\nu > 0$. Now you should prove that the distribution P_X is approximately equal to one where we ignore all the atypical sequences,

$$Q_{\vec{X}}(\vec{x}) = \begin{cases} P_{\vec{X}}(\vec{x})/P_{\vec{X}}[\mathcal{G}_\nu] & \text{if } \vec{x} \in \mathcal{G}_\nu, \\ 0 & \text{if } \vec{x} \notin \mathcal{G}_\nu. \end{cases} \tag{4}$$

So prove that the trace distance between the original distribution P_X and the "truncated" Q_X vanishes,

$$\lim_{n\to\infty} \delta(P_{\vec{X}}, Q_{\vec{X}}) = 0. \tag{5}$$

Now use that result to prove that, for fixed ϵ and ν,

$$\lim_{n\to\infty} \frac{1}{n} H^\epsilon_{\min}(\vec{X}) \geq H(X) - \nu. \tag{6}$$

You know (page 15 of the script) that $H_{\min}(X) \leq H(X) \leq H_{\max}(X)$ and should get an equality from there. The max entropy can also be shown to be the same as the Shannon entropy in the i.i.d. limit, but you don't have to show that in the exercise.

Exercise 3.3 Quantum-Telepathy Game: Introduction

Welcome back to the quantum world! In the past few weeks we talked about fundamental concepts of information theory. Now we will spend two weeks in a quick recap of quantum mechanicss basics (chapter 4.1 of the script, which, by the way, you will be assumed to know by heart in the lectures), before we get to the fun stuff—quantum information. However, we would like to give you a taste of how quantum systems can be more powerful than classical bits in information-related tasks. Let us begin by introducing an *entangled* state.
The entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is one of those intrinsically odd things about quantum mechanics: it is a pure state of two systems, such that if you measure it in the $\{|0\rangle, |1\rangle\}$ basis on each side the outcomes will be random but *always the same* in both systems. Check that by yourself: apply the joint measurement with outcomes $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ to that state and see the probabilities of obtaining each outcome (page 30 of the script).
We will study entanglement in detail later on, and see everything one can do using entangled states (things like teleportation and superdense coding). For now, look at the state that Alice and Bob share in the exercise, $\frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle)$. This, as you may guess, is another entangled state. Remember that the $\{|+\rangle, |-\rangle\}$ basis is just a rotation of the computational basis for a qubit, $\{|0\rangle, |1\rangle\}$ (we will get back to this next week). In this case, however, they will always obtain opposite outcomes if they each measure their qubit in a very obvious basis. Solve this exercise and let's jump to 3.2 *b*).
Now we are going to see what happens to a quantum state (in particular, an entangled quantum state) when one measures only one part of it. Alice, Bob and Charlie share the state $\frac{1}{\sqrt{2}}(|000\rangle - |111\rangle)$. Charlie will try to measure his qubit in a special basis so that Alice and Bob recover the very convenient state $\frac{1}{\sqrt{2}}(|+-\rangle + |-+\rangle)$ if he gets outcome b_0 and a similar state (just in a different basis) when he gets b_1. Try measuring the state in a couple of bases and see what happens. Use the postulate of page 30 to get the post-measurement state, and partial trace (pages 25–26) to trace out Charlie's final qubit and see what Alice and Bob get. Don't cheat!
Once you found the right measurement, you know something quite useful: Charlie is able to leave Alice and Bob with one of two states that they can use to obtain different bits x_1 and x_2. He cannot predict which of those two states it will be, since the outcomes of his measurement are random, but after his measurement he can tell them which state they have by sending them a single bit, b. It's time to tackle the general game.

Exercise 3.4 Quantum-Telepathy Game: The Full Story

The game starts with n collaborating players $P_1, P_2, \ldots, P_n$ who each have a qubit of a large state $|\Psi\rangle$ in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$. In other words, player P_i has control of the qubit in the space $\mathcal{H}_i$. Then two of them will be randomly selected and separated from the other players. These two players, let's label them P_1 and P_2, are separated without the knowledge of which other player was selected, and they cannot communicate with any of the players, including each other. The remaining $n-2$ players are allowed to communicate with each other, and do a measurement on the qubits they each control. They can then send a bit b (either 0 or 1) to the two separated players. P_1 and P_2 output bits x_1 and x_2 respectively. They win the game if $x_1 \neq x_2$.

To get a better feeling for the exercise let us look at what players can do in the classical (meaning non-quantum) case. Well, essentially they can divide the players in four equally sized groups such that they only lose the game if the two chosen players belong to the same group, *i.e.* they have a 75% chance of winning. Check Fig. 3 for details.

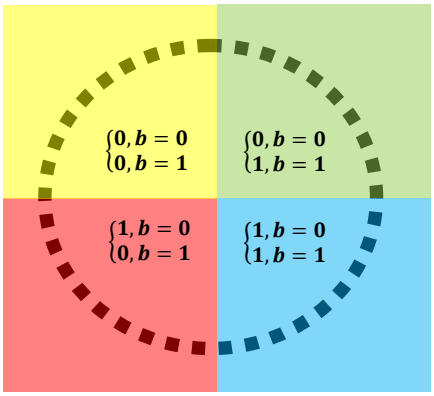

Figure 3: The best strategy in the classical case: players are divided in four groups of the same size. Each group has different instructions about what bit to output depending on the input bit. For instance if a member of the green group ends up in the dark room, she should output 0 if the bit the other players pass her is a 0, and 1 if the input is a 1. The players left in the room know which bit to send to the confined players according to their groups. For instance if one of them is yellow and the other is green, they should send bit 1 so that they output different bits. They will only lose the game if both players belong to the same group, with 25% of probability.

In the quantum version of this game players are allowed to share a quantum state and to perform local measurements. However players are not allowed to communicate the results of their measurements to the caged ones, except for that single bit. Of course if the total state is entangled then measuring a part of it will produce some changes in the rest, but it is not as if they could send some useful information to the other players that way, because they cannot control the outcomes of non deterministic measurements.

There is a strategy, however, that allows players to always win the game. They need to create a very special $n-$qubit state in $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n$ and distribute it, leaving player i in control of $\mathcal{H}_i$. Then the players that are left in the room will measure their qubits one by one (uin the same basis as Charlie did) and remove them from the total system.

The magic here is that since they made a very clever choice of initial state (one where all the qubits are entangled in that neat way you have in the exercise sheet), applying those operators will change the shared state in a nice, controlled way. Of course that they cannot predict to which state it will collapse before performing the measurement, but they can keep track of the state of the system by checking their measurement results.

The next step of their strategy is to send one bit of information to the two caged players. What do you think they should tell them? And what can the prisoners do with their state to ensure they will output different bits?