MAPSS, Les Diablerets
14.07.26

INTRODUCTION TO QUANTUM INFORMATION THEORY

- Ramona Wolf -

Lecture I: Classical Information Theory

What is information?

→ How much do you learn from an event?

Example: Coin toss

- Possible outcomes: heads, tails
- Probability for each outcome: $\frac{1}{2}$ (ideal coin)

Suppose you flip the coin once and see "heads". How much did you learn?

⇝ Theory of information

1.1 Probabilities

A central concept are random variables. This is a mathematical formalisation of an object or a process that depends on random (= non-deterministic) events.

It consists of a triple $(X, \mathcal{X}, P_X)$, where

X - random variable

$\mathcal{X}$ - alphabet (= possible values $x \in \mathcal{X}$ that X can attain, also called realisations)

P_X - probability distribution over $\mathcal{X}$, i.e., a function

$$P_X : \mathcal{X} \to [0,1] \text{ such that } \sum_{x \in \mathcal{X}} P_X(x) = 1.$$

What happens if we have two random variables (e.g., two coins)?

$\Rightarrow$ $(X, \mathcal{X}, P_X)$ and $(Y, \mathcal{Y}, P_Y)$

If we throw both together, we can calculate the probability distribution over all possible pairs: "Joint probability distribution"

$$P_{XY}(x,y) = P_X(x) \cdot P_Y(y) \quad \text{if the two dice are independent!}$$

If they are not independent (maybe we connected them somehow), the equation above does not hold. Then we cannot use P_X and P_Y to obtain P_{XY} directly, but we need to know more about the correlation between X and Y.

It can happen that we have side-information about the die that might change the distribution of the outcomes.

Example: Assume someone hands you a die, and it can either be a fair one (as described above) or a biased one which always shows a 6.

You don't know which die you get, both possibilities are equally likely. This is described by a random variable Y, $\mathcal{Y} = \{\text{"fair"}, \text{"biased"}\}$, with

$$P_Y(\text{"fair"}) = \frac{1}{2}, \quad P_Y(\text{"biased"}) = \frac{1}{2}.$$

As soon as you learn the value of Y, i.e., which die you got, you know more about the likelihood of the outcomes. This is captured by conditional probabilities:

$$P_{X|Y}(x|\text{"fair"}) = \frac{1}{6} \quad \forall x \in \mathcal{X}, \qquad P_{X|Y}(x|\text{"biased"}) = \begin{cases} 1 & \text{if } x = 6 \\ 0 & \text{else} \end{cases}$$

Some useful relations between different probability distributions:

$$P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)}$$

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x) \cdot P_X(x)}{P_Y(y)} \qquad \text{(Bayes' rule)}$$

Notation: We often write simply $p(x,y)$ if it is clear what distribution we are talking about.

1.2 Entropies and Information

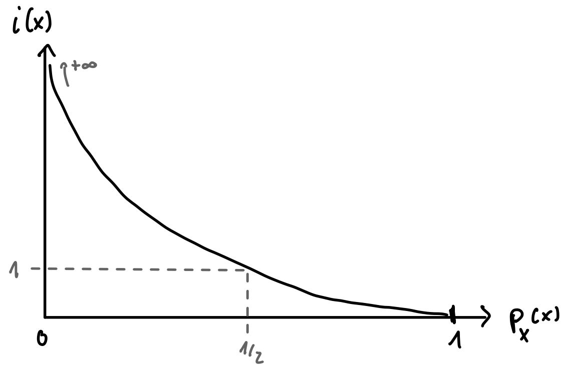
Why are we sometimes more surprised than other times?

Example: (i) Fair die vs.
(ii) a die where 5 sides show 1 and 1 side shows 2.

Suppose both dice show a 2. We are more surprised to see it at the biased die (ii) than the fair die (i).

$\Rightarrow$ Seeing less likely events is more surprising.

Definition: The information content of a realisation x of X is defined as
$$i(x) = -\log(p_X(x)).$$



Why is this a good choice?, It has desirable mathematical properties:

1. It only depends on the probability of the realisation, not on the label.
2. It is continuous in p.
3. It is monotonically decreasing in p and non-negative.
4. It is additive: (for independent realisations)
$$\begin{aligned} i(x_1, x_2) &= -\log(p(x_1, x_2)) \\ &= -\log(p(x_1)\,p(x_2)) \\ &= -\log(p(x_1)) - \log(p(x_2)) \\ &= i(x_1) + i(x_2) \end{aligned}$$

How do we capture the general amount of information a random variable X possesses?

→ Expectation value of $i(x)$ over all realisations $x \in \mathcal{X}$.

Definition: The Shannon entropy of a discrete random variable X with outcome set $\mathcal{X}$ and probability distribution $p(x)$ is
$$H(X) = -\sum_{x\in\mathcal{X}} p(x) \log(p(x)).$$

Here, the logarithm is to base 2, so the resulting number quantifies the information in <u>bits</u>.

Sometimes people also use the natural logarithm, but the two choices only differ by a constant:

$$\log_e(x) = \frac{\log_2(x)}{\log_2(e)} .$$

What if $p(x) = 0$? We use the convention that $0 \cdot \log 0 = 0$, which is usually done in information theory. Also backed up by limit $\lim_{x \to 0} x \log x = 0$.

Example: Fair coin X, $\mathcal{X} = \{\text{heads}, \text{tails}\}$, $p(x) = \frac{1}{2}\ \forall x$

$$H(X) = -\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1$$

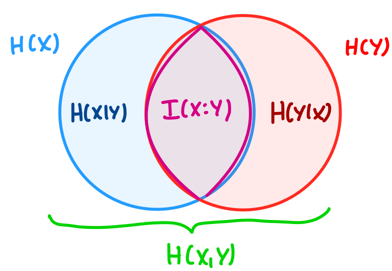
$\longrightarrow$ You get 1 bit of information when learning about the outcome of a coin toss $\to$ 1 bit $\equiv$ 2 possibilities, equally likely

Based on the definition of Shannon entropy, one can define other entropies:

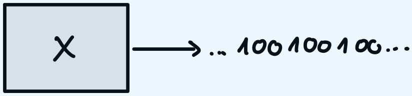
Definition: Let $(X, \mathcal{X}, P_X)$ and $(Y, \mathcal{Y}, P_Y)$ be two random variables and $P_{XY}(x,y)$ the joint probability distribution. We define:

- joint entropy: $H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$
- conditional entropy: $H(X|Y) = -\sum_{x,y} p(x,y) \log \underbrace{\frac{p(x,y)}{p(y)}}_{= p(x|y)} = H(X,Y) - H(Y)$
- mutual information: $I(X:Y) = H(X) + H(Y) - H(X,Y)$

Relationship between entropies:



__Example:__ Suppose we have a source with periodic output.



We are supposed to guess the next output of the source, with varying side information. How is our uncertainty quantified?

1. We have not seen any output. Then the probability distribution is given by the frequency of the respective symbol in the sequence:

$$p(x=0) = \frac{2}{3} \quad , \quad p(x=1) = \frac{1}{3} \qquad (*)$$

$$\Rightarrow H(X) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} \approx 0{,}918$$

2. One known symbol: Now we have some side information, so we have to compute $H(X|Y)$. Y is the bit we have seen.

Let us rewrite it as

$$H(X|Y) = -\sum_{x,y} p(x,y)\log p(x|y)$$
$$= -\sum_{x,y} p(y)\,p(x|y)\log p(x|y)$$

$p(y)$ is simply given by $(*)$

next bit

$p(x\|y)$	$x=0$	$x=1$
$y=0$	$1/2$	$1/2$
already know $y=1$	1	0

$$\Rightarrow H(X|Y) = \frac{2}{3} \approx 0{,}66$$

our uncertainty has decreased!

3. Two known symbols: If we already know two consecutive symbols, we should be certain about what comes next.

$p(x\|y)$	$x=0$	$x=1$
$y=00$	0	1
$y=01$	1	0
$y=10$	1	0

$$\Rightarrow H(X|Y) = 0.$$

1.3 Communication theory

One can define a variety of entropy measures. What makes some of them so particularly useful is their operational interpretation, i.e., what types of questions they allow us to answer

Typically, messages contain a lot of redundancy, which means they can be compressed but are still readable. For example:

GD DS NT PLY DC

I compressed the message by removing its vowels.*

Shannon's noiseless coding theorem tells us how far a message can be compressed while still being able to recover it.
The key quantity is the entropy:

Suppose X describes a source that outputs data $x \in \mathcal{X}$ with probability $p(x)$. Then, there exists a compression scheme that allows the information to be stored using $H(X)$ bits per symbol on average such that the information can be recovered perfectly. Any compression scheme that uses fewer than $H(X)$ bits/symbol will inevitably loose information.

Example: source X, $\mathcal{X} = \{a, b, c, d\}$, $p(x) = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$

Without any data compression, you need 2 bits per symbol to encode the outputs:

a	b	c	d
00	01	10	11

But the Shannon entropy is

$$H(X) = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{8}\log\frac{1}{8} + \frac{1}{8}\log\frac{1}{8}\right) = \frac{7}{4} < 2$$

So we can compress the data to use less bits per output:
Idea: More likely symbols get shorter codewords.

a	b	c	d
0	10	110	111

(*)

* If you struggle to read it, this is the full message: GOD DOES NOT PLAY DICE

How many bits do you need on average to encode a symbol?

$$\sum_x p(x) \cdot \text{\# bits of codeword for } x = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = \frac{7}{4},$$

which is $= H(X)$

This is the best we can hope for. What happens if we try to compress it further?
For example,

a	b	c	d
0	1	01	10

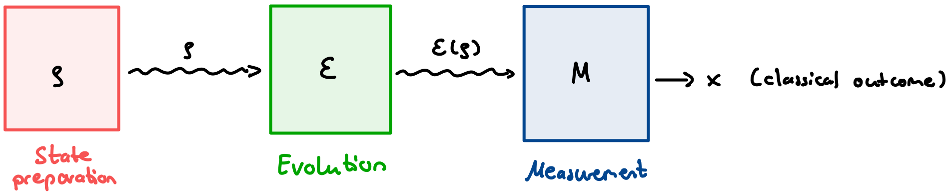
This uses only 1,5 bits per symbol. But we cannot reliably retrieve the information from the codeword:

0 | 0 | 1 | 1 0 | 1 | 0 | 0 = acdda
= aabdbaa

Using ⊛ there is only one possibility: 0 | 0 | 1 1 0 | 1 0 | 0 = aacba

Lecture II: The Formalism of Quantum Information

General structure of a quantum mechanical experiment:



2.1 Quantum states

Quantum states, in general, are vectors in a Hilbert space $\mathcal{H}$ (over $\mathbb{C}$)

Bra-ket notation:

- We denote a vector in $\mathcal{H}$ as $|\psi\rangle$ ("ket")
- The dual space of $\mathcal{H}$ is $\mathcal{H}^*$, the space of linear map from $\mathcal{H}$ to $\mathbb{C}$ (because we consider $\mathcal{H}$ over $\mathbb{C}$). We denote a vector in $\mathcal{H}^*$ as $\langle\varphi|$ ("bra")
- Inner product: $(\varphi, \psi) = \langle\varphi|\psi\rangle$

Some important concepts:

- Given two d-dimensional vectors $|\psi_1\rangle = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix}$, $|\psi_2\rangle = \begin{pmatrix} b_1 \\ \vdots \\ b_d \end{pmatrix}$ their <u>inner product</u> is given by $\langle\psi_1|\psi_2\rangle$.

  How to calculate this? Note that $\langle\psi_1| = (a_1^* \dots a_d^*)$. Then

  $$\langle\psi_1|\psi_2\rangle = (a_1^* \dots a_d^*) \begin{pmatrix} b_1 \\ \vdots \\ b_d \end{pmatrix} = \sum_{i=1}^{d} a_i^* b_i .$$

- The <u>norm</u> of a ket vector $|\psi\rangle = \begin{pmatrix} a_1 \\ \vdots \\ a_d \end{pmatrix}$ is defined as

  $$\| \, |\psi\rangle \, \| = \sqrt{\langle\psi|\psi\rangle} = \sqrt{\sum_{i=1}^{d} a_i^* a_i} = \sqrt{\sum_{i=1}^{d} |a_i|^2} .$$

  If $\| \, |\psi\rangle \, \| = 1$, we say that $|\psi\rangle$ is normalised.

- An <u>orthonormal basis</u> of a Hilbert space $\mathcal{H}$ is a set of labelled vectors $\{|e_i\rangle\}_i$ such that

  $$\langle e_i | e_j \rangle = \delta_{ij} \quad \forall \, i,j$$

  and every vector $|\psi\rangle \in \mathcal{H}$ can be written as a linear combination of the basis vectors:

  $$|\psi\rangle = \sum_i \langle e_i|\psi\rangle \, |e_i\rangle .$$

Qubits – The most important example of quantum states (for us)

Qubits are two-level quantum systems that generalise classical bits.

Hilbert space: $\mathcal{H}_{\text{Qubit}} = \mathbb{C}^2$

Possible basis: $\begin{pmatrix}1\\0\end{pmatrix} \equiv |0\rangle$ ($\hat{=}$ classical "0")

$\begin{pmatrix}0\\1\end{pmatrix} \equiv |1\rangle$ ($\hat{=}$ classical "1")

Computational basis

(Z basis)

While classical bits can <u>only</u> be in the state $|0\rangle$ or $|1\rangle$, qubits can attain every state in $\mathbb{C}^2$, i.e., arbitrary linear combinations of $|0\rangle$ and $|1\rangle$.

General state:
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$
with $\alpha, \beta \in \mathbb{C}$ s.t. $|\alpha|^2 + |\beta|^2 = 1 \Rightarrow \| |\psi\rangle \| = 1$

The corresponding "bra" vector is $\langle\psi| = \alpha^* \langle 0| + \beta^* \langle 1|$.

There are many basis choices (infinitely many, in fact) for a Hilbert space.

For example:
$$|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \quad |-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

Hadamard basis

(X basis)

It is related to the computational basis:
$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

<u>Quiz:</u>

1. Is $|\psi\rangle = \frac{1}{4}|0\rangle + \frac{2}{5}|1\rangle$ a valid qubit state?
2. For which values is $|\psi_\theta\rangle = \cos\theta\,|0\rangle + \sin\theta\,|1\rangle$ a valid qubit state?

Solution: 1. No: $\left(\frac{1}{4}\right)^2 + \left(\frac{3}{8}\right)^2 = \frac{1}{16} + \frac{9}{64} = \frac{4}{64} + \frac{9}{64} = \frac{13}{64} \neq 1$.

2. For all θ: $\cos^2\theta + \sin^2\theta = 1$

What if we don't know perfectly which state the quantum system is in? Let's say, we know the system is in one of the states $|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle$ with respective probabilities $p_1, p_2, \dots, p_n$.

(with $p_i \geq 0$, $\sum_i p_i = 1$)

This is called an <u>ensemble</u>: $\boxed{\{p_i, |\psi_i\rangle\}_{i=1,\dots,n}}$.

The state of the system is then written as a density matrix $\rho \in \mathcal{B}(\mathcal{H})$

bounded linear operators on $\mathcal{H}$

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|.$$

This is not a vector in $\mathcal{H}$ anymore, but an operator on $\mathcal{H}$.

It is formally defined as follows:

Definition: A density matrix ρ is an operator on $\mathcal{H}$ such that

1. It is normalised: $Tr(\rho) = 1$.
2. It is Hermitian: $\rho^\dagger = \rho$.
3. It is positive semi-definite: $\rho \geq 0$, i.e., $\langle \varphi | \rho | \varphi \rangle \geq 0 \quad \forall\, |\varphi\rangle \in \mathcal{H}$.

We denote the set of all density operators on $\mathcal{H}$ as $S(\mathcal{H})$.
(= bounded linear operators on $\mathcal{H}$ that fulfil some extra conditions)

Why do we have two different ways of expressing quantum states?

- Density matrices capture our uncertainty about the exact state of the system
  $\rightarrow$ classical randomness, could also be due to noise
- In modern QT, density matrices are the more natural objects (follow from derivations via algebras)
- Historically: Closed systems, captured by pure states. Schrödinger equation But pure states are very restrictive (no noise, no open quantum systems)

2.2 Evolution of quantum systems

What do you know about the evolution of quantum systems?
$\rightarrow$ Schrödinger equation; only works for closed systems

Here: More general

What mathematical object describes the evolution of a quantum system?
What do we need?
$\rightarrow$ A map that maps quantum states to quantum states

Definition: A quantum channel is a linear, completely positive, trace-preserving map.

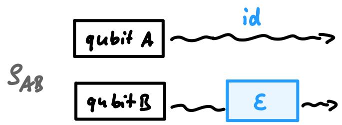
- Linear: $\mathcal{E}(\alpha\rho + \beta\sigma) = \alpha\, \mathcal{E}(\rho) + \beta\, \mathcal{E}(\sigma)$
- Trace-preserving: $Tr(\rho) = Tr(\mathcal{E}(\rho)) \quad \forall \rho \in S(\mathcal{H})$
  (density operators are mapped to density operators)

- Completely positive: Start with <u>positivity</u>:

  A linear map $\mathcal{E}: \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is positive if $\mathcal{E}(\rho) \geq 0 \quad \forall\, \rho \geq 0$
  (positive semi-definite)

  Why is this not sufficient?

  Consider the following situation:



$(id \otimes \mathcal{E})\,(\rho_{AB})$

- ρ_{AB} is a valid 2-qubit state.
- I do nothing on qubit A, and apply a positive map on qubit B.
  Is the resulting state $(id \otimes \mathcal{E})\,(\rho_{AB})$ a valid density matrix?

  $\rightarrow$ No! (Not necessarily) Example: $\mathcal{E}$ is the transpose on B

Positivity does not extend to this scenario. We need to require <u>complete positivity</u>:

A linear map $\mathcal{E}: \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is completely positive if the map

$$\mathcal{E} \otimes id_n : \mathcal{B}(\mathcal{H}_A) \otimes \mathcal{B}(\mathbb{C}^n) \longrightarrow \mathcal{B}(\mathcal{H}_B) \otimes \mathcal{B}(\mathbb{C}^n) \qquad (id_n = \text{identity map on } \mathbb{C}^n)$$

is positive for all $n \in \mathbb{N}$.

2.3 Measurements

What should a measurement describe?

Input: quantum state ρ; output: probability distribution $p_\rho(i)$ over possible outcomes i
(depends on ρ)

<u>Definition:</u> A Positive Operator-Valued Measure (POVM) is a set $\{M_i\}_i$ of operators that are

1. positive: $\forall i: \; M_i \geq 0$ $\qquad p_\rho(i) \geq 0 \quad \forall \rho$
2. Hermitian: $\forall i: \; M_i^\dagger = M_i$ $\qquad p_\rho(i) \in \mathbb{R} \quad \forall \rho$
3. Complete: $\sum_i M_i = \mathbb{1}$ $\qquad \sum_i p_\rho(i) = 1 \quad \forall \rho$

What properties do we want to know when talking about measurements?

1. What is the probability of outcome x?
2. In what state is the quantum system after the measurement?

Regarding outcome probabilities, these can be computed as follows:

- Pure state: $\Pr[x] = \langle\psi| M_x |\psi\rangle$
- Mixed state: $\Pr[x] = \mathrm{Tr}(M_x \rho) \underset{(\text{if } \rho \text{ pure})}{=} \mathrm{Tr}(M_x |\psi\rangle\langle\psi|) = \mathrm{Tr}(\langle\psi| M_x |\psi\rangle) = \langle\psi| M_x |\psi\rangle.$

Special case: Projective measurement

Additional property: $M_i M_j = \delta_{ij} M_i$ (orthogonal measurements)

Then we can also write down the post-measurement state:

- Pure state: $|\psi'\rangle = \dfrac{M_x |\psi\rangle}{\sqrt{\langle\psi| M_x |\psi\rangle}}$
- Mixed state: $\rho' = \dfrac{M_x \rho M_x}{\mathrm{Tr}(M_x \rho)}$

Measurement in a basis

Let $\{|u_i\rangle\}_i$ be an ONB on a Hilbert space $\mathcal{H}$ and let $|\psi\rangle \in \mathcal{H}$. A measurement in the basis $\{|u_i\rangle\}_i$ corresponds to the POVM where $M_i := |u_i\rangle\langle u_i|$.

Check properties:

- positive: $M_i \geq 0 \Rightarrow \langle\varphi|u_i\rangle\langle u_i|\varphi\rangle = |\langle\varphi|u_i\rangle|^2 \geq 0$
- Hermitian: $M_i^\dagger = (|u_i\rangle\langle u_i|)^\dagger = |u_i\rangle\langle u_i| = M_i$
- complete: $\sum_i M_i = \sum_i |u_i\rangle\langle u_i| = \mathbb{1}$

It is even projective:

$$M_i M_j = |u_i\rangle\underbrace{\langle u_i|u_j\rangle}_{=\delta_{ij}}\langle u_j| = \delta_{ij} |u_i\rangle\langle u_i|$$

Two important examples:

1. Z-basis $\{|0\rangle, |1\rangle\}$, $M_0 = |0\rangle\langle 0|$, $M_1 = |1\rangle\langle 1|$

   General state: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$

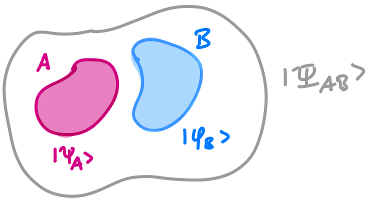
   $\Pr[0] = \langle\psi| M_0 |\psi\rangle = (\alpha^* \langle 0| + \beta^* \langle 1|)\, |0\rangle\langle 0|\, (\alpha|0\rangle + \beta|1\rangle) = \underbrace{\alpha^*\alpha}_{=|\alpha|^2} \overbrace{\langle 0|0\rangle}^{=1}$

   $\Pr[1] = \langle\psi| M_1 |\psi\rangle = |\beta|^2$

2. X-basis $\{|+\rangle, |-\rangle\}$

   $M_+ = |+\rangle\langle +|$, $M_- = |-\rangle\langle -|$

2.4 Composite systems



Composite system: Systems are composed via the tensor product

$|\psi\rangle_A \in \mathcal{H}_A \,, \quad |\varphi\rangle_B \in \mathcal{H}_B$

$\Rightarrow |\psi_A\rangle \otimes |\varphi\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B$

Basis: $\{ |\psi_i\rangle_A \}_i \,, \quad \{ |\varphi_j\rangle_B \}_j$

$\Rightarrow$ Basis of the composite system: $\{ |\psi_i\rangle_A \otimes |\varphi_j\rangle_B \}_{i,j}$

What are the dimensions? ($\hat{=}$ # basis states)

$\dim(\mathcal{H}_A) = d_A \,, \quad \dim(\mathcal{H}_B) = d_B \quad \Rightarrow \quad \dim(\mathcal{H}_A \otimes \mathcal{H}_B) = d_A \cdot d_B$

<u>Example:</u> Two qubits

$\mathcal{H}_A = \mathbb{C}^2, \quad \mathcal{H}_B = \mathbb{C}^2$

Basis for both: $\{ |0\rangle, |1\rangle \} \,, \quad |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Basis for $\mathcal{H}_A \otimes \mathcal{H}_B$: $\{ |0\rangle \otimes |0\rangle \,, |0\rangle \otimes |1\rangle \,, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle \}$

$= \{ |00\rangle, |01\rangle, |10\rangle, |11\rangle \}$

$|00\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \,, \quad |01\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad |10\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad |11\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

$\Rightarrow$ General vector in the 2-qubit space:

$$|\psi\rangle_{AB} = \alpha \cdot |00\rangle + \beta \cdot |01\rangle + \gamma \cdot |10\rangle + \delta \cdot |11\rangle = \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{pmatrix}$$

For density matrices, it works the same:

$\rho_A \in S(\mathcal{H}_A), \quad \rho_B \in S(\mathcal{H}_B) \Rightarrow \quad \rho_A \otimes \rho_B \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$

<u>The interesting part:</u> There are more states in $\mathcal{H}_A \otimes \mathcal{H}_B$ than just product states of the form $|\psi\rangle_A \otimes |\varphi\rangle_B$! (we will get to that in the next session)

For example: $$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Is it possible to write this as $|\varphi\rangle_A \otimes |\eta\rangle_B$? → No!
The state is <u>entangled</u>.

For pure states, this is defined as follows:

<u>Definition:</u> A pure bipartite state $|\psi\rangle_{AB}$ is called entangled if it cannot be written as a product state $|\varphi\rangle_A \otimes |\eta\rangle_B$ for any choice of states $|\varphi\rangle_A$ and $|\eta\rangle_B$. Otherwise it is called separable.

And for mixed states:

<u>Definition:</u> A mixed bipartite state ρ_{AB} is called entangled if it cannot be written as a linear convex combination of product states, i.e., it can <u>not</u> be written as

$$\rho_{AB} = \sum_x p_x \, \varphi_A^x \otimes \eta_B^x$$

for some probability distribution $\{p_x\}$ and density matrices $\sigma_A^x \in S(\mathcal{H}_A)$, $\eta_A^x \in S(\mathcal{H}_B)$.

But if $\rho_{AB} \neq \rho_A \otimes \rho_B$, what can we say about the marginal states ρ_A and ρ_B ?
We can use the <u>partial trace</u> for this:

<u>Definition:</u> Let $\rho_{AB} \in S(\mathcal{H}_A \otimes \mathcal{H}_B)$ and let $\{|i\rangle_B\}_i$ be an orthonormal basis for $\mathcal{H}_B$. The partial trace over B is then defined as

$$\mathrm{Tr}_B(\rho_{AB}) = \sum_i (\mathbb{I}_A \otimes \langle i|_B)\, \rho_{AB}\, (\mathbb{I}_A \otimes |i\rangle_B).$$

$$\equiv \sum_i \langle i|\rho_{AB}|i\rangle_B .$$

Example: Parity measurement of two qubits

(Even parity: $|00\rangle$ or $|11\rangle$; Odd parity: $|01\rangle$ or $|10\rangle$)

1. First possibility: Measure in the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$

$M_{00} = |00\rangle\langle 00|$, $M_{01} = |01\rangle\langle 01|$, $M_{10} = |10\rangle\langle 10|$, $M_{11} = |11\rangle\langle 11|$

Probability for outcome "even":

$$q_{even} = q_{00} + q_{11} = \mathrm{Tr}(M_{00}\rho) + \mathrm{Tr}(M_{11}\rho) = \mathrm{Tr}(|00\rangle\langle 00|\rho) + \mathrm{Tr}(|11\rangle\langle 11|\rho)$$
$$= \langle 00|\rho|00\rangle + \langle 11|\rho|11\rangle$$

Post-measurement state: Mixture of two even-parity outcome post-meas. states:

$$\rho_{even} = \frac{1}{q_{even}} \left(M_{00}\rho M_{00} + M_{11}\rho M_{11}\right)$$
$$= \frac{1}{q_{even}} \left(|00\rangle\langle 00|\rho|00\rangle\langle 00| + |11\rangle\langle 11|\rho|11\rangle\langle 11|\right)$$
$$= \frac{1}{q_{even}} \left[(\langle 00|\rho|00\rangle)|00\rangle\langle 00| + (\langle 11|\rho|11\rangle)|11\rangle\langle 11|\right]$$

2. Second possibility: We directly use projectors for the relevant subspaces without measuring the qubits individually.

$M_{even} = |00\rangle\langle 00| + |11\rangle\langle 11|$, $M_{odd} = |01\rangle\langle 01| + |10\rangle\langle 10| = \mathbb{1} - M_{even}$

Probability for outcome "even" is unchanged:

$$q'_{even} = \mathrm{Tr}(M_{even}\rho) = \langle 00|\rho|00\rangle + \langle 11|\rho|11\rangle = q_{even}$$

(Would be weird if it was different; then the two measurements wouldn't measure the same property)

But the post-measurement state is different:

$$\rho'_{even} = \frac{1}{q'_{even}} M_{even}\rho M_{even}$$

To see that there is actually a difference, consider the state

$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$:

Clearly, $q_{even} = q'_{even} = 1$. But the post-measurement states are:

$\rho_{even} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11| = \frac{1}{2}\mathbb{I}$ $\Rightarrow$ All information was destroyed

$\rho'_{even} = |\Phi^+\rangle\langle\Phi^+|$ $\Rightarrow$ The state is unchanged!

Same information, different effect on the state.

Lecture III: Quantum entropy and applications

3.1 Quantifying quantum information

Now we want to generalise the concept of entropies to quantum systems, i.e., we want to quantify the information resp our uncertainty about the state of a quantum system:

Definition: Let A be a quantum system prepared in a state $\rho_A \in \mathcal{B}(\mathcal{H}_A)$. The von Neumann entropy of A is defined as

$$H(A)_\rho = -\operatorname{Tr}(\rho_A \log \rho_A).$$

This formula has a useful relation to the eigenvalues λ_i of the state ρ_A:

Let $\rho_A = \sum_i \underbrace{\lambda_i}_{\text{eigenvalues}} \underbrace{|e_i\rangle\langle e_i|_A}_{\text{eigenvectors}}$, then $H(A) = -\sum_i \lambda_i \log \lambda_i$.

Where does our uncertainty about a quantum system come from?

Examples: 1) $\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\Rightarrow$ eigenvalues $\{\frac{1}{2}, \frac{1}{2}\}$

$\Rightarrow H(A) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$ (same as the Shannon entropy of a fair coin)

2) $\rho_A = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $\Rightarrow$ eigenvalue $\{1\}$ (pure state)

$\Rightarrow H(A) = -1 \cdot \log 1 = 0$ $\Rightarrow$ no uncertainty in which quantum state the system is, since ρ_A is a pure state!

ρ_A describes our knowledge of the system A. If it is in a pure state, we know exactly in which state it is. A mixed state captures that we only know the state of the system with a certain probability.

$$\rho_A = \underbrace{\frac{1}{2}}_{\text{With 50\% the system is in state } |0\rangle}|0\rangle\langle 0| + \underbrace{\frac{1}{2}}_{\text{With 50\% the system is in state } |1\rangle}|1\rangle\langle 1|$$

Captures a limitation of our knowledge

This is different than a state in superposition:

$$|\psi\rangle_A = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

This system is not in either state with 50% probability. Only when measured, it is projected onto one of the states.

<u>Important property!</u> $H(A)=0 \Leftrightarrow \rho_A$ is a pure state.

"$\Rightarrow$": $H(A) = -\sum_i \lambda_i \log \lambda_i$, λ_i eigenvalues of ρ_A

If $H(A)=0$, then $\lambda_i \log \lambda_i = 0 \ \forall i$.

This is fulfilled either if $\lambda_i = 0$ or $\lambda_i = 1$

Since $\sum_i \lambda_i = 1$, there is exactly one label i^* for which $\lambda_{i^*} = 1$.

$\Rightarrow \rho_A$ has exactly one eigenvalue $=1$ (the rest $=0$)

$\Rightarrow \rho_A = |\psi\rangle\langle\psi|_A$ for some vector $|\psi\rangle_A$.

"$\Leftarrow$": ρ_A pure $\Rightarrow \rho_A = |\psi\rangle\langle\psi|_A \Rightarrow$ eigenvalues $\{1\} \Rightarrow H(A) = -1 \cdot \log 1 = 0$

<u>**Definition:**</u> For a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$ in a state $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ the joint quantum entropy is given by

$$H(AB)_\rho = -\mathrm{Tr}\, \rho_{AB} \log \rho_{AB}.$$

<u>**Definition:**</u> Let $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a bipartite quantum state. The conditional entropy is given by

$$H(A|B)_\rho = H(AB)_\rho - H(B)_\rho.$$

Let us study these entropies a bit more.
Note that in the classical case, all entropies are positive – which makes sense considering that they all capture how much information we learn from a random variable.

In the quantum case, this can be different, making the interpretation of these quantities more challenging.

The von Neuman entropy $H(A)$ of a quantum state ρ_A is always positive: $H(A) \geq 0$.

One can also show that for any separable state $\rho_{AB} = \sum_x p_x \, \sigma_A^x \otimes \eta_B^x$, the conditional entropy is non-negative:

$$H(A|B) \geq 0.$$

But what happens for entangled states?
Consider the state

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}} \left(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B \right) = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)$$

Let us calculate the conditional entropy $H(A|B) = H(AB) - H(B)$:

- $H(AB)$: The system AB is in a pure state, so there is no uncertainty about the state of the joint system

  $\Rightarrow$ The entropy of a pure state is always 0 $\Rightarrow$ $H(AB) = 0$

- $H(B)$: To calculate this, we need the marginal state of system B :

$$\rho_B = Tr_A\left(|\Phi^+\rangle\langle\Phi^+|_{AB}\right)$$

(use: $\langle 0|0\rangle = \langle 1|1\rangle = 1$
$\langle 0|1\rangle = \langle 1|0\rangle = 0$)

$$= \frac{1}{2}\Big[{}_B\langle 0| \left(|00\rangle\langle 00| + \cancel{|00\rangle\langle 11|} + \cancel{|11\rangle\langle 00|} + \cancel{|11\rangle\langle 11|}\right)|0\rangle_B$$
$$+ \langle 1| \left(\cancel{|00\rangle\langle 00|} + \cancel{|00\rangle\langle 11|} + \cancel{|11\rangle\langle 00|} + |11\rangle\langle 11|\right)|1\rangle_B\Big]$$

$$= \frac{1}{2}\Big[(\mathbb{1}_A \otimes \langle 0|_B)(|0\rangle_A \otimes |0\rangle_B)(\langle 0|_A \otimes \langle 0|_B)(\mathbb{1}_A \otimes |0\rangle_B)$$
$$+ (\mathbb{1}_A \otimes \langle 0|_B)(|0\rangle_A \otimes |0\rangle_B)(\langle 1|_A \otimes \langle 1|_B)(\mathbb{1}_A \otimes |0\rangle_B) + \ldots\Big]$$

$$= \frac{1}{2}\Big[|0\rangle\langle 0|_A \underbrace{\langle 0|0\rangle}_{=1}\langle 0|0\rangle_B + |0\rangle\langle 1|_A \langle 0|0\rangle\underbrace{\langle 1|0\rangle}_{=0}{}_B + \ldots \Big]$$

$$= \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right)$$

$\Rightarrow$ $H(B) = 1$ (example above)

$\Rightarrow$ $H(A|B) = H(AB) - H(B) = -1$ this is negative !

What does this mean ??

$H(AB) = 0$ $\Rightarrow$ There is nothing left to learn about the state of the joint system, it is perfectly known.

$H(B) = 1$ $\Rightarrow$ This is the maximum value the entropy can take; You know nothing about the individual system.
($= H(A)$)

How can we know more about the joint system AB than its parts A and B individually?

Schrödinger [1935]: " I would not call that one but rather the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought. "

3.2 Application: Quantum Teleportation

This strange observation is maybe best understood when analysing an application. For this, we need to introduce some operators.

The Pauli operators

These operators form a basis for the operators on $\mathbb{C}^2$:

$$X = |0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad X|0\rangle = |1\rangle, \quad X|1\rangle = |0\rangle \qquad \text{"bit flip"}$$

$$Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Y|0\rangle = i|1\rangle, \quad Y|1\rangle = -i|0\rangle$$

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad Z|0\rangle = |0\rangle, \quad Z|1\rangle = -|1\rangle$$

To obtain a basis for $\mathbb{C}^2$, we also have to add the identity matrix:

$$\mathbb{1} = |0\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Note that all Pauli operators square to the identity: $X^2 = Y^2 = Z^2 = \mathbb{1}$

With this we can introduce the Bell basis:

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

$$|\Phi^-\rangle_{AB} = (\mathbb{1}_A \otimes Z_B)|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes Z|0\rangle_B + |1\rangle_A \otimes Z|1\rangle_B)$$
$$= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$

$$|\Psi^+\rangle_{AB} = (\mathbb{1}_A \otimes X_B)|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$

$$|\Psi^-\rangle_{AB} = (\mathbb{1}_A \otimes (-iY))|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

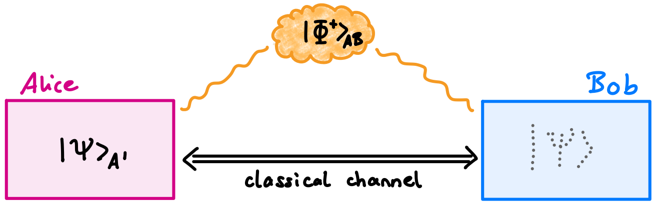
$\Rightarrow$ Basis on the 2-qubit space.
All basis vectors are related by local operations on Bob's system.

Note: Since this is a basis, one can also construct a 2-qubit POVM from the Bell states where the output is two bits $(00, 01, 10, 11)$ via

$$M_{00} = |\Phi^+\rangle\langle\Phi^+|, \quad M_{01} = |\Phi^-\rangle\langle\Phi^-|, \quad M_{10} = |\Psi^+\rangle\langle\Psi^+|, \quad M_{11} = |\Psi^-\rangle\langle\Psi^-|$$

With these states, Alice and Bob can carry out a protocol called quantum teleportation.

The setup:



Goal: Alice wants to transmit a qubit state $|\psi\rangle_A$, but she only has acces to a classical channel.

Can she achieve her goal with only the classical channel? No!
What if she had more copies of the state?

Recall: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ (general qubit state)

and $\Pr[0] = |\alpha|^2$, $\Pr[1] = |\beta|^2$ when measuring in the computational basis $\{|0\rangle, |1\rangle\}$

So with many copies, Alice could measure them and collect statistics for $|\alpha|^2$ and $|\beta|^2$ (never mind she also needs the relative phase between α and β).

But here, Alice only has a single copy of the state she wants to send to Bob.

What if Alice and Bob already share an entangled state, for example $|\Phi^+\rangle_{AB}$
Then they can run the following protocol:

1. Alice takes the qubit $|\psi\rangle_{A'}$ she wants to send and her part of $|\Phi^+\rangle_{AB}$ and measures the two qubits in the Bell basis, producing a two-bit outcome.

What happens to Bob's part of $|\Phi^+\rangle_{AB}$ when Alice sees the outcome 00?

$$(M_{00}^{A'A} \otimes \mathbb{1}_B)\, |\psi\rangle_{A'} |\Phi^+\rangle_{AB}$$

$$= |\Phi^+\rangle_{A'A} \langle\Phi^+| \psi\rangle_{A'} |\Phi^+\rangle_{AB}$$

$$= |\Phi^+\rangle_{A'A} \langle\Phi^+| \left[(\alpha|0\rangle_{A'} + \beta|1\rangle_{A'}) \tfrac{1}{\sqrt{2}} (|00\rangle_{AB} + |11\rangle_{AB}) \right]$$

$$= |\Phi^+\rangle_{A'A} \tfrac{1}{\sqrt{2}} (\langle 00| + \langle 11|)_{A'A} \left[\tfrac{1}{\sqrt{2}}\, \alpha|000\rangle + \alpha|011\rangle + \beta|100\rangle + \beta|111\rangle \right]_{A'AB}$$

$$= |\Phi^+\rangle_{A'A} \tfrac{1}{\sqrt{2}} \left(\tfrac{1}{\sqrt{2}} \alpha \langle 00|00\rangle_{A'A} |0\rangle_B + \beta \langle 11|11\rangle_{A'A} |1\rangle_B \right)$$

$$= |\Phi^+\rangle_{A'A} \tfrac{1}{2} (\alpha|0\rangle_B + \beta|1\rangle_B) = \tfrac{1}{2} |\Phi^+\rangle_{AB} |\psi\rangle_B$$

exactly the state Alice wanted to send!

What is the probability that the outcome is 00?

$\Rightarrow \Pr[00] = \langle\psi|_{A'} \mathrm{Tr}_B(\langle\Phi^+|_{AB})\, M_{00}\, |\psi\rangle_{A'} \mathrm{Tr}_B(|\Phi^+\rangle_{AB}) = \frac{1}{4}$

What if Alice gets a different outcome?

Suppose she gets the outcome 01. Note that

$$(M_{01}^{A'A} \otimes \mathbb{1}_B)\, |\psi\rangle_{A'} |\Phi^+\rangle_{AB} = |\Phi^-\rangle\langle\Phi^-|_{A'A}\, |\psi\rangle_{A'} |\Phi^+\rangle_{AB}$$

$$= |\Phi^-\rangle_{A'A} \langle\Phi^+|(\mathbb{1}_{A'} \otimes Z_A)\, |\psi\rangle_{A'} |\Phi^+\rangle_{AB}$$

$$= |\Phi^-\rangle_{A'A} \otimes Z_B |\psi\rangle_B$$

(by repeating the calculation from above)

So instead of $|\psi\rangle$, Bob's state is $Z|\psi\rangle$.

He can then apply an operation to get the original state:

$ZZ|\psi\rangle = \mathbb{1}\,|\psi\rangle = |\psi\rangle$.

Analogously, for the other outcomes:

$10 \rightarrow X|\psi\rangle$

$11 \rightarrow -iY|\psi\rangle$

If Bob knows Alice's outcome, in each case he can apply the corresponding Pauli operator to his system to reconstruct the state $|\psi\rangle$.

So the protocol needs a second step:

2. Alice communicates the outcome (2 bits) to Bob. Bob then performs the corresponding correction to his quantum state.

In summary: By communicating 2 bits of classical information and using 1 shared entangled 2-qubit state as a resource, Alice can send an arbitrary qubit state to Bob.

How does this connect to the negative conditional entropy?

The protocol above requires an entangled state as a resource.

If $H(A|B) > 0$, there is no entanglement between Alice and Bob. A negative conditional entropy $H(A|B) < 0$ indicates entanglement and therefore witnesses the presence of the resource required to implement the quantum teleportation protocol.

Therefore, a negative conditional entropy can be interpreted as <u>the potential for future quantum communication</u>.