AQI: Advanced Quantum Information
Lecture 6 (Module 2): Distinguishing Quantum States

January 28, 2013

Lecturer: Dr. Mark Tame

Introduction

With the emergence of new types of information, in this case quantum states, we enlarge the class of dynamical processes beyond those considered in classical information theory. A nice example of this is the problem of distinguishing quantum states. Classically, it's possible to distinguish between two states, for instance the letters 'a' and 'b' with perfect certainty, in principle. However, quantum mechanically it's not always possible to distinguish between two quantum states. In particular, non-orthogonal quantum states can't be reliably distinguished and this phenomena lies at the heart of a variety of tasks in quantum information processing. Here, quantum states containing hidden information not accessible to measurements play a key role in quantum algorithms and quantum cryptography.

1 Distinguishing orthonormal states

- Consider two parties, Alice and Bob. Alice chooses a state $|\psi_i\rangle$ $(1 \leq i \leq n)$ from a fixed set of states known to both parties.
- Alice gives Bob the state $|\psi_i\rangle$ and he must identify the index i of the state Alice has given him.
- Bob uses the following set of measurement operators $\{M_i\}_{i=0}^n$, where $M_i = |\psi_i\rangle\langle\psi_i|$ and $M_0 = \mathbb{1} - \sum_{i\neq 0} |\psi_i\rangle\langle\psi_i|$, so that $\sum_m M_m^\dagger M_m = \mathbb{1}$.
- If the state $|\psi_i\rangle$ was prepared by Alice, then the probability that measurement outcome i occurs is given by $p(i) = \langle\psi_i| M_i |\psi_i\rangle = 1$, so that outcome i occurs with certainty and one can reliably distinguish orthonormal states.

Note: Here I'm talking about a 'single shot' measurement. In other words, a single system is prepared and a single measurement is made.

General rules: $|\psi\rangle \rightarrow \frac{M_i|\psi\rangle}{\sqrt{\langle\psi|M_i^\dagger M_i|\psi\rangle}}$, where $p(i) = \langle\psi| M_i^\dagger M_i |\psi\rangle = \langle\psi| M_i |\psi\rangle$ as $M_i^\dagger M_i = M_i$ for projective measurements. For a measurement set $\{E_i\}$ we have the corresponding operators $M_i = \sqrt{E_i}$, with $E_i = M_i^\dagger M_i = M_i$ for projective and E_i otherwise. In general, $\rho \rightarrow \frac{M_i \rho M_i^\dagger}{\mathrm{Tr}(M_i \rho M_i^\dagger)}$.

2 Distinguishing non-orthonormal states

- Consider Alice prepares one of two non-orthonormal states $|\psi_1\rangle$ and $|\psi_2\rangle$.
- Bob uses the set of measurement operators $\{M_j\}$ and depending on the outcome of his measurement he tries to guess the index i of Alice's state using some rule $i = f(j)$, *e.g.* $f(1) = 1$, $f(2) = 2$ and $f(3) = 2$.

Now, let's assume that Bob <u>can</u> distinguish Alice's non-orthonormal states.

- If Alice prepares $|\psi_1\rangle$ then the probability of Bob obtaining outcome j such that $f(j) = 1$ is 1.
- If Alice prepares $|\psi_2\rangle$ then the probability of Bob obtaining outcome j such that $f(j) = 2$ is 1.
- Using the definition $E_i = \sum_{j:f(j)=i} M_j^\dagger M_j$ we have that

$$\begin{aligned} \langle\psi_1| E_1 |\psi_1\rangle &= 1 \\ \langle\psi_2| E_2 |\psi_2\rangle &= 1 \end{aligned} \tag{1}$$

  Here, $\sum_j M_j^\dagger M_j = \mathbb{1} \rightarrow \sum_i E_i = \mathbb{1}$ so that $\sum_i \langle\psi_1| E_i |\psi_1\rangle = 1$ and since $\langle\psi_1| E_1 |\psi_1\rangle = 1$ we have $\langle\psi_1| E_2 |\psi_1\rangle = 0 \rightarrow \sqrt{E_2}\,|\psi_1\rangle = 0$.

In the above, Eq. (1) is our assumption of distinguishability. We'll now see that this assumption leads to a contradiction.

- Decompose $|\psi_2\rangle = \alpha\,|\psi_1\rangle + \beta\,|\varphi\rangle$, where $|\varphi\rangle$ is orthonormal to $|\psi_1\rangle$, with $|\alpha|^2 + |\beta|^2 = 1$ and $|\beta| < 1$ (as $|\psi_1\rangle$ and $|\psi_2\rangle$ are not orthogonal).
- Based on our assumption we have $\sqrt{E_2}\,|\psi_1\rangle = 0$ so that $\sqrt{E_2}\,|\psi_2\rangle = \beta\sqrt{E_2}\,|\varphi\rangle$ and

$$\langle\psi_2| E_2 |\psi_2\rangle = |\beta|^2 \langle\varphi| E_2 |\varphi\rangle \leq |\beta|^2 < 1 \tag{2}$$

  where we have used $\langle\varphi| E_2 |\varphi\rangle \leq \sum_i \langle\varphi| E_i |\varphi\rangle = \langle\varphi|\varphi\rangle = 1$.

Therefore our assumption contradicts a basic property of the non-orthonormality of the states ($|\beta| < 1$) and can't be true! Thus, one can't reliably distinguish non-orthonormal states.

Note: While perfect reliability isn't possible, it is possible to distinguish non-orthonormal states some of the time. You'll see this in the problem sheet, where POVM formalism is used for the measurements.

3 Entropy and Mutual Information

In general the amount by which quantum states can be perfectly distinguished is quantified by the Holevo bound. This is an upper bound on the accessible information that can be obtained by measuring a quantum state and thus is an upper bound on the measure of how well Bob can infer the state that Alice prepared. Before we introduce the Holevo bound we first need to know about some concepts in quantum information.

3.1 Shannon Entropy (classical)

The Shannon entropy associated with the probability distribution $\{p_1, \ldots, p_n\}$ for possible values of a symbol X is defined as

$$H(X) \equiv H(p_1, \ldots, p_n) \equiv -\sum_x p_x \log_2 p_x \tag{3}$$

with $0\log_2 0 \equiv 0$ ($\lim_{x\to 0} x\log x = 0$). Entropy is measured in 'bits', although you may see a different base for the logarithm used sometimes, *e.g.* $\log_b p_x$ with $b = 2$ as bits, $b = e$ as nats and $b = 10$ as dits. The Shannon entropy quantifies the resources needed to store information.

Let's see an example:

- A source produces four symbols 'a'$= 00$, 'b'$= 01$, 'c'$= 10$ and 'd'$= 11$. Without compression 2 bits are consumed for each use of the source.
- However, suppose that 'a' is produced from the source with probability $p_a = 1/2$, 'b' with $p_b = 1/4$, 'c' with $p_c = 1/8$ and 'd' with $p_d = 1/8$. We can use this bias to compress the data from the source so that fewer bits are used for more commonly occurring symbols.
- Consider the compression scheme: 'a'$= 0$,'b'$= 10$,'c'$= 110$ and 'd'$= 111$. A compressed message from the source contains on average $H(X) = \frac{1}{2}\times 1 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = 7/4$ bits of information per use of the source. This is the Shannon entropy.

The Shannon entropy can also be thought of as quantifying how much information we gain on average when we learn the value of a symbol X produced by a source. Alternatively, one can think of it as measuring the uncertainty about X before we learn the value. We are certain X contains on average $7/4$ bits of information and we are uncertain how much more information it contains.

3.2 Joint Entropy (classical)

The joint entropy is defined as

$$H(X, Y) \equiv -\sum_{x,y} p(x, y)\log_2 p(x, y) \tag{4}$$

and measures the total uncertainty about the pair of symbols (X, Y).

3.3 Conditional Entropy (classical)

If we know the value of Y we have acquired $H(Y)$ bits of information about the pair (X, Y). The remaining uncertainty about (X, Y) is associated with our remaining lack of knowledge about X, given that we know Y. The entropy of X conditional on knowing Y is defined as

$$H(X|Y) \equiv H(X, Y) - H(Y). \tag{5}$$

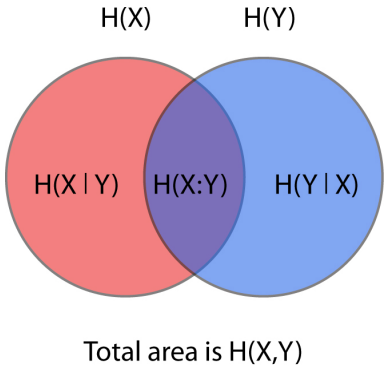
3.4 Mutual Information (classical)

This measures how much information X and Y have in common. If we add the information content of X, $H(X)$, to that of Y, $H(Y)$, then information common to both X and Y is counted twice, while information

not common is counted only once. Subtracting off the joint information $H(X,Y)$ gives us the common or mutual information of X and Y

$$H(X:Y) \equiv H(X) + H(Y) - H(X,Y). \tag{6}$$

Sometimes the symbol I is also used for the mutual information. Using Eq. (5), we also have $H(X:Y) = H(X) - H(X|Y)$. The entropy Venn diagram below is helpful



3.5 Data Processing Inequality (classical)

Let $X \to Y \to Z$ be a Markov chain (a sequence of random variables such that Z is independent of X given Y). We then have

$$H(X) \geq H(X:Y) \geq H(X:Z), \tag{7}$$

with the first inequality saturated iff, given Y, it's possible to reconstruct X. This inequality tells us that if we have a random variable, X, that is subject to noise producing Y, then further actions by us (data processing) can't be used to increase the amount of mutual information between the output of the process, Z, and the original information X. In other words, once information is lost it's gone forever!

3.6 Von Neumann Entropy (quantum)

The von Neumann entropy of a quantum state ρ is the generalisation of the Shannon entropy for a symbol X and is defined as

$$S(\rho) \equiv -\mathrm{Tr}(\rho \log_2 \rho) \tag{8}$$

If λ_x are the eigenvalues of ρ, *i.e.* $\rho = \sum_x \lambda_x |\psi_x\rangle\langle\psi_x|$ we can write

$$S(\rho) = -\sum_x \lambda_x \log_2 \lambda_x. \tag{9}$$

The von Neumann entropy is measured in bits (as we use $\log_b$, with $b = 2$). Also, $S(\rho) : 0 \to 1$ for qubits and $S(\rho) : 0 \to \log_2 d$ for qudits.

3.7 Joint Entropy (quantum)

The joint entropy is defined as

$$S(A,B) \equiv -\mathrm{Tr}(\rho^{AB} \log_2 \rho^{AB}), \tag{10}$$

where ρ^{AB} is the density matrix of system AB.

3.8 Conditional Entropy (quantum)

The conditional entropy is defined as

$$S(A|B) \equiv S(A,B) - S(B). \tag{11}$$

3.9 Mutual Information (quantum)

The mutual information is defined as

$$\begin{aligned} S(A:B) &\equiv S(A) + S(B) - S(A,B) && (12)\\ &= S(A) - S(A|B) \\ &= S(B) - S(B|A). && (13) \end{aligned}$$

4 The Holevo Bound

Ok, let's put all these things together to show the maximum amount of information that can be extracted from a quantum system with a single measurement. We'll be mixing classical and quantum information theory concepts for this.

- Let Alice have a classical source producing symbols $X = 0, \ldots, n$ according to a probability distribution $p_0, \ldots, p_n$. Bob's goal is to determine the value of X as best he can.
- To do this Alice prepares a quantum state ρ_X chosen from a fixed set of non-orthogonal states $\rho_0, \ldots, \rho_n$ with the same probability distribution as X, and gives it to Bob who makes a measurement and tries to make the best guess he can about X, based on his measurement outcome Y.
- A measure of how much information Bob has gained about X through his measurement is the mutual information $H(X:Y)$ between X and the measurement outcome Y, *i.e.* how much information X and Y have in common. Note, this isn't the same as the conditional entropy $H(X|Y)$.
- From the data processing inequality we know that Bob can infer X from Y iff $H(X:Y) = H(X)$ and that in general $H(X:Y) \leq H(X)$. Bob's best strategy is therefore to choose a measurement that maximises $H(X:Y)$, bringing it as close to $H(X)$ as he can.
- Thus, we define Bob's 'accessible information' to be the maximum of the mutual information $H(X:Y)$ over all possible measurement strategies. This accessible information is a measure of how well Bob can do at inferring the state that Alice prepared.

The Holevo bound is

$$H(X:Y) \leq S(\rho) - \sum_x p_x S(\rho_x), \tag{14}$$

where $\rho = \sum_x p_x \rho_x$. This bound is an upper bound on the accessible information for a single-shot measurement. The right hand side of Eq. (14) is called the Holevo χ quantity, $\chi = S(\rho) - \sum_x p_x S(\rho_x)$. One can show that if $\rho = \sum_x p_x \rho_x$, where p_x are a set of probabilities and the ρ_x are density operators, then

$$S(\rho) \leq \sum_x p_x S(\rho_x) + H(X) \tag{15}$$

with equality iff ρ_x have support on orthogonal subspaces.

In the present example, we have non-orthogonal states prepared by Alice, so

$$H(X) > S(\rho) - \sum_x p_x S(\rho_x), \tag{16}$$

using Eq. (14) we get

$$H(X) > H(X:Y). \tag{17}$$

So it's impossible for Bob to determine X with perfect reliability based on his measurement result Y. This generalises our understanding of the problem of distinguishing non-orthogonal states using quantum information theory techniques.

4.1 Example

- Alice prepares the state $|0\rangle$ with probability $1/2$ and the state $\cos\theta\,|0\rangle + \sin\theta\,|1\rangle$ with probability $1/2$. Thus, she sends the state

$$\rho = \rho_1 + \rho_2 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}, \tag{18}$$

  to Bob, which has eigenvalues $\lambda_\pm = \frac{1}{2}(1 \pm \cos\theta)$.

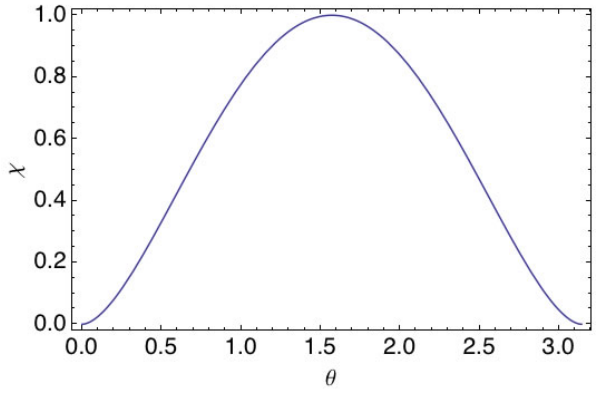
- We use the Holevo χ quantity, $\chi = S(\rho) - \sum_x p_x S(\rho_x)$ to find the upper bound on the amount of information that Bob can extract about the state Alice prepared. Here we have

$$\begin{aligned} S(\rho) &= -\lambda_+ \log_2 \lambda_+ - \lambda_- \log_2 \lambda_- & (19) \\ S(\rho_1) &= -1 \log_2 1 = 0 \\ S(\rho_2) &= -1 \log_2 1 = 0. & (20) \end{aligned}$$

  Thus,

$$H(X:Y) \leq -\lambda_+ \log_2 \lambda_+ - \lambda_- \log_2 \lambda_- = \chi \tag{21}$$

The figure below shows the behaviour of the Holevo bound χ. Note: When $\theta = \pi/2$, Alice produces states from an orthogonal set and Bob (for some measurement strategy he must try to find) can determine with certainty the states, as $H(X:Y) \leq 1$.



References

- M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2000).
- J. Preskill, *Quantum Information lecture notes*, http://www.theory.caltech.edu/people/preskill/ph229/#lecture (2004).