Classical capacities of a qubit

Aidan J. Klobuchar
BSM FALL SEMESTER 2010

1 Introduction

In Classical Information theory the smallest piece of information is the classical bit, which can take the values of either 0 or 1. That is, a classical bit has only two pure states. Two bits have 4 pure states, 3 bits have $2^3 = 8$ pure states and so on.

When talking about a quantum bit or **qubit**, one considers a two-level quantum system: a quantum system which is described by a Hilbert space $\mathcal{H}$ of dimension two (e.g. the spin-content of a spin 1/2 particle). The state space of a quantum system with Hilbert space $\mathcal{H}$ can be identified with the space of **density operators** $S_1^+(\mathcal{H})$; that is, operators ρ acting on $\mathcal{H}$ having the properties

$$\rho \geq 0, \quad \mathrm{Tr}(\rho) = 1. \tag{1}$$

When $\dim(\mathcal{H}) = 2$, the convex body $S_1^+(\mathcal{H})$ is precisely a 3-dimensional ball and thus each of its border points is extremal. (In general — when $\dim(\mathcal{H}) > 2$ — the shape of $S_1^+(\mathcal{H})$ is much less understood, and cannot be so simply described as in the 2-dimensional case. It will not be simply a ball and not all of its border points will be extremal. Nevertheless, it will still have continuously many extremal points.) So in contrast to a classical bit which has only 2 pure states, a qubit has infinitely many. However, this does not neccessarily mean that we can store more (classical) information in a qubit than in a classical one. The point is that though our qubit has infinitely many different pure states, it is impossible to distinguish these states *with certanity*. This is a fundamental fact, and cannot be circumvented by some better measuring device.

As was seen in class, in case of a two-level system, one can distinguish with certanity between at most 2 states. So in this respect a single qubit performs very like a classical one. However, this does not make a single qubit and a single classical bit necessarily equivalent. Perhaps it is possible to distinguish between $n > 2$ states of the qubit not with certainty, but in a

way that is — in some sense — "closer" to certanity than what we can have with a classical bit.

How to define 'closer to certanity' is an issue the following sections will describe. In general, we may view our qubit as a memory — or as it is often called in the literature: a channel — in which there is a ingoing and an outgoing information (someone chooses a certain state from a preveously fixed set of states and puts the qubit into the selected state, then passes it to another person who will have to try to determine: in which state the electron is). So one may investigate the issue from the point of view some kind of a *channel capacity*. As we shall see, various capacity-like quantities of a qubit coincides with that of a classical bit.

2 A capacity concept based on a game

To model the amount of information a single classical or quantum bit can carry, we consider a channel which is realized by passing a single bit from a sender, Alice, to a receiver, Bob. We assume that the bit, when passed to Bob, is in an independent state[1] from rest of the world in reach of Bob.

One can then introduce various different quantities all trying to reflect the capacity of this channel to send useful information. Here, instead of the usually considered *Shannon capacity*, in order to familiarize with the concepts, we shall start with a somewhat simpler quantity. We shall introduce this quantity by considering a game involving a single bit.

Suppose a \$1 bill is put randomly and with equal probability into one of n boxes. Bob must pick one of the boxes and he gets what is inside that box. Now, to Alice it is revealed under which box the \$1 bill is. However, Alice cannot directly tell this to Bob (in which case Bob could always get the \$1 bill with certanity). Instead, she is only allowed to send to Bob a classical or a quantum bit whose state she can manipulate as she wishes. (That is, she is allowed to send a classical or a quantum bit of information.)

They may agree on some scheme beforehand. For example, played with a classical bit, Alice and Bob can agree that the bit-value 0 will mean that the money is in box number 1 and the bit-value 1 will mean that the money is not in that box. Alternatively, they may agree that the bit-value 0 will

[1] In case of dense-coding this condition does not hold: the sent particle is entangled with another one which is with Bob. Infact, in that case it *is* possible to transmit with a qubit more than one bit of classical information

mean that the money is in boxes 1 - $\lfloor \frac{n}{2} \rfloor$ and the bit-value 1 will mean that the money is in boxes $\lfloor \frac{n}{2} + 1 \rfloor$ - n.

The question then becomes, what is the expected value of the money Bob may win? This expected value (after a certain normalization) will be our measure of capacity.

In the classical case, the answer is straightforward. It is not too difficult to see that the ideal strategy is to have the measured value 0 correspond to half of the boxes and the measured value of 1 correspond to the other half of the boxes. Then with n defined as the number of boxes and $E[\$]$ as the expected value of the money won, we obtain

$$E[\$] = \text{ prob. of finding the bill} = \frac{1}{n/2} = \frac{2}{n}. \tag{2}$$

In the above, we have argued by using "common sense". To make things more rigorous and also to be able to proceed to more complex arguments, let us try now to formalize the basic concepts.

The chosen encoding scheme followed by Alice can be described by a $n \times 2$ matrix A. This matrix contains the probability values of Alice putting her bit into a particular state upon seeing that the money is in a particular box. That is, $A_{i,j}$ is the probability that upon seeing the money in the i-th box, Alice will send the bit to Bob in state $j \in \{0, 1\}$. Naturally, all entry values need to be between 0 and 1 (since they are probability values) and the sum of the elements in each row must be 1 (given, that the money is in a particular box, Alice will *either* put her bit into state 0 *or* state 1: the sum of the respective probabilities must be 1).

The chosen decoding scheme followed by Bob can be described by a $2 \times n$ matrix B. This matrix contains the probability values of Bob picking a particular box upon receiving the bit from Alice in a particular state. That is, $B_{j,k}$ is the probability that upon receiving the bit in state $j \in \{0, 1\}$, Bob will pick box number k. Likewise to Alice's encoding matrix, also B must have its entry values between 0 and 1 and must have its rows sum to 1.

For the $n = 3$ case (there are 3 boxes) these encoding and decoding "tables" (matrices) would like this:
Let us now talk about the probability $p_{i \to k}$; that is, the probability of Bob picking box k given that the \$1 bill is under box k. Of course, to obtain its value, we must take account of both the encoding and the decoding proba-

Table 1: Alice's encoding

	bit = 0	bit = 1
box nr 1	$p^A_{1\to 0}$	$p^A_{1\to 1}$
box nr 2	$p^A_{2\to 0}$	$p^A_{2\to 1}$
box nr 3	$p^A_{3\to 0}$	$p^A_{3\to 1}$

Table 2: Bob's decoding

	box nr 1	box nr 2	box nr 3
bit = 0	$p^B_{0\to 1}$	$p^B_{0\to 2}$	$p^B_{0\to 3}$
bit = 1	$p^B_{1\to 1}$	$p^B_{1\to 2}$	$p^B_{1\to 3}$

bilities:

$$p_{i\to k} = (p^A_{i\to 0})(p^B_{0\to k}) + (p^A_{i\to 1})(p^B_{1\to k}). \tag{3}$$

That is, the values $\{p_{i\to k}\}$ are obtained by multiplying Alice's encoding matrix with Bob's decoding matrix through standard matrix multiplication. We shall refer to the obtained matrix T of transitional probabilities as the **channel table** of the certain encoding-decoding scheme. This is an $n \times n$ matrix with nonnegative entries in which every row adds to 1.

Now, back on the topic of the maximum amount of money that can be won in the money game, it is clear that the expected value of the money won is given by

$$E(\$) = \frac{1}{n}\sum_j p_{j\to j} = \frac{1}{n}\mathrm{Tr}(T) = \frac{1}{n}\mathrm{Tr}(AB), \tag{4}$$

where T is the channel table, A is the encoding and B is the decoding matrix. Now, as every entry of A is less or equal than one and every entry of B is nonnegative, we have that

$$Tr(AB) = \sum_{k,l} A_{k,l}B_{l,k} \leq \sum_{k,l} B_{l,k} = 2, \tag{5}$$

as the sum of each row of B must be 1 and there are 2 rows in B. Thus

$$E[\$] = \frac{1}{n}Tr(AB) \leq \frac{2}{n}. \tag{6}$$

So classically, the money we can expect to obtain when a single \$1 bill is placed randomly into one of n boxes is at most $\$\frac{2}{n}$. The question then becomes, can this value be improved quantumly? We will answer this question in the following section.

3 The "money bound" on the qubit channel

Let us discuss the problem in a quantum framework using the Hilbert space $\mathcal{H}$ of the qubit. After learning the location of the money, Alice will put her qubit into a state given by a density operator. If the money is in box j, Alice will put her qubit into state $\rho_j \in S_1^+(\mathcal{H})$. That is, encoding is nothing else than a map $\{1, 2, \ldots, m\} \to S_1^+(\mathcal{H})$. Note that in the quantum description it seems we have avoided of considering probabilities in the choices of Alice. However, it only seems so: we did not assume ρ_j to be *pure*.

The qubit will be then passed to Bob's measuring device. Informally, this is a device with n lights, where the incoming qubit will trigger one of the lights to go off and thus signal to Bob which box to choose.

Formally, such a device is described by a POVM $(E_1, \ldots E_n)$; i.e. collection of positive operators summing to I. The channel table containing the transitional probabilities is nothing else than the matrix $(\mathrm{Tr}(\rho_j E_k))_{\{j,k\}}$. What can we say about the amount of money that can be won with a qubit in our money game?

The spectrum of a density operator ρ is always contained in the interval $[0, 1]$. Hence $\rho \leq I$ and $I - \rho$ is a positive operator and so if E is another positive operator then

$$\mathrm{Tr}((I - \rho)E) \geq 0 \;\Leftrightarrow\; \mathrm{Tr}(\rho E) \leq \mathrm{Tr}(E). \tag{7}$$

Thus all elements in the k-th column of the channel table are smaller or equal than $\mathrm{Tr}(E_k)$ and so in particular the expected value of the money won is smaller or equal than

$$\frac{1}{n}(\mathrm{Tr}(E_1) + \mathrm{Tr}(E_2) + \ldots + Tr(E_n)) = \frac{1}{n}\mathrm{Tr}(I) = \frac{1}{n}\dim(\mathcal{H}) = \frac{2}{n}, \tag{8}$$

since the dimension of the Hilbert space of a qubit is 2. Thus, no matter what is the actual measuring device used by Bob, and what is the encoding scheme used by Alice, a single quantum bit can make win no more money in our little game than a classical bit. This amount of money that can be

won is a form of channel capacity, as it gives an indication of the amount of information a bit may hold.

4 Shannon Channel Capacity

It would be interesting to see if a single classical and quantum bit share the same maximum capacity in the sense of the *Shannon Channel Capacity*. To look at a quantum system as classical channel, we need to fix an *encoding*; that is we need to fix a map $i \mapsto \rho_i$ from letter of the input alphabet $\{a, b, ...\}$ to the set of density operators $S_1^+(\mathcal{H})$ (i.e. to the set of states of our quantum system). Decoding, from the mathematical point, is a convex structure preserving map from $S_1^+(\mathcal{H})$ to the output alphabet $\{\alpha, \beta, \ldots\}$ and as was discussed, is given by a POVM $\{E\}$. (From the physical point of view decoding is the actual device chosen by Bob, which picks up the sent quantum system and after examining it produces an output letter. To take account of a certain device, one then needs to specify how do the probabilities of the outcoming letters depend on the incoming state of the system; this is why we are considering decoding as the discussed map.) In our money game example,the input and output alphabets were identical, though this does not need to be the case.

The **Shannon channel capacity** is simply the maximum[2] amount of *mutual information* $I(\pi : \tilde{\pi})$ between the *coding probability distribution* $\{\pi\}$ and the probability distribution $\tilde{\pi}$ of the outcoming letter. Here

- the **coding probability distribution** is the list of probabilities that Alice will code a particular letter of the input alphabet (for example, the probability of Alice coding a is given by the value π_a — i.e. π_a describes how often a appears in Alice's messages)

- the **outcome probability distribution** $\tilde{\pi}$ is the list of probabilities that Bob will decode a particular leter of the output alphabet (for example, the probability of Bob will finally decode α is given by the value π_α.

The outcome probability distribution is determined by the *transitional probabilities* and the coding probability distribution π. The **transitional prob-**

[2]In the finite case the existence of a maximum can be easily shown. In general however, one should be more careful and consider a *supremum* instead of a maximum.

ability $p_{i\to j}$ is the probability that if the input is set to i the output will be j. With a fixed coding $\{\rho\}$ and decoding $\{E\}$, as was discussed

$$p_{i\to j} = Tr(\rho_i E_j). \tag{9}$$

That is, what we called a channel table is merely the collection of these values. Knowing π and the values $\{p_{i\to j}|i,j\}$ the outcome distribution can be calculated as

$$\tilde{\pi}_j = \sum_i \eta_{i,j} \tag{10}$$

Here $\{\eta_{i,j}|i,j\}$ is the *joint distribution* of the income and outcome:

$$\eta_{i,j} = \pi_i p_{i\to j} \tag{11}$$

is the probability that Alice will encode i and Bob will receive j. The mutual information $I(\pi : \tilde{\pi})$ is then

$$I(\pi : \tilde{\pi}) = H(\pi) + H(\tilde{\pi}) - H(\eta) \tag{12}$$

where $H(X)$ is the **entropy** of a probability distribution $X = (x_1, \ldots, x_n)$:

$$H(X) = -\sum_k x_k \log(x_k) \tag{13}$$

where the logarithms are traditionally taken base 2 (so that a single classical bit would turn out to have a channel capacity of 1 unit). Using that a probability distribution always adds to 1, and using the properties of the log function, by substitution one arrives to the following well-known formula:

$$I(\pi : \tilde{\pi}) = \sum_{i,j} \pi_i p_{i\to j} \log\left(\frac{p_{i\to j}}{\sum_k \pi_k p_{k\to j}}\right). \tag{14}$$

(Here i runs over the input alphabet, that is, the 'letters' which Alice can code in, and j runs over the output alphabet, or the different 'letters' which Bob's measuring device can read out.)

Now, suppose our channel relies on an n-level quantum system (that is, our density operators $\rho_1, \rho_2, \ldots$ and POVM are given on an n-dimensional

Hilbert space). In this case then, what is the maximum value that the (classical) Shannon capacity C of the channel may be? By for example [1, Thm. 2.1] we have that

$$C \leq \sup_{\pi}\{H(\sum_{k}\pi_k\rho_k) - \sum_{k}\pi_k H(\rho_k)\} \tag{15}$$

where the supremum is taken over all probability distributions $\{\pi\}$ and $H(X) = \mathrm{Tr}(h(X))$ is the **von Neumann entropy** of a density operator X. Here h is the entropy function

$$h(x) = \begin{cases} -x\log(x), & \text{if } \mathrm{x} > 0 \\ 0, & \text{if } x = 0 \end{cases} \tag{16}$$

and $h(X)$ is defined *via* the spectral calculus. In other words, $H(X)$ is the (classical) entropy of the distribution of eigenvalues (taken with multiplicities) of the density operator X.

Since the von Neumann entropy of a density operator is nonnegative, we further have that

$$C \leq \sup_{\pi} H(\sum_{k}\pi_k\rho_k). \tag{17}$$

For any probability distribution $\{\pi\}$, the convex combination $\sum_k \pi_k\rho_k$ is a density operator. So we can further estimate the capacity by taking a supremum over the set of *all* density operators and hence

$$C \leq \sup_{\rho} H(\rho) = H((\frac{1}{n})I) = \log(n). \tag{18}$$

(It is well known that the entropy of a probability distribution is maximal if the distribution is uniform. That is, the highest von Neumann entropy is achieved when all eigenvalues of the density operator coincide; that is, when the density operator is a multiple of the identity.)

This upper bound indicates that the maximum channel capacity of a quantum channel is no greater than the maximum value of a classical channel, which is $\log(n)$. Note that the upper bound of $\log(n)$, on the other hand, is achievable. Indeed, let $\rho_1, \ldots, \rho_n$ be n 1-dimensional orthogonal projections summing to the identity. Then setting $E_j := \rho_j$ $(j = 1, \ldots, n)$ we have that $\{E_j\}_j$ is a POVM (actually it is more specifically a PVM: a projection

valued measure). Using our choice of density operators and POVM, the channel table we obtain is simply the $n \times n$ identity matrix, since we have

$$\mathrm{Tr}(\rho_i E_j) = \mathrm{Tr}(\rho_i \rho_j) = \mathrm{Tr}(\delta_{i,j} \rho_j) = \delta_{i,j}. \tag{19}$$

Then further setting π to be the uniform distribution $(1/n, 1/n, \ldots, 1/n)$ we get that with our choices $I(\pi, I) = \log(n)$. Since the capacity C is obtained as a supremum, this shows that $C \geq \log(n)$. Together with the upper bound (18) this shows that in this case C is precisely $\log(n)$.

Note that the Shannon Channel Capacity and the capacity defined in the previous section ("Money Capacity") reflect different ideas and it is easy to find cases where two schemes (channel tables) can have an equivalent Money / Shannon capacity and have a differing Shannon / Money capacity, respectively. Regardless, they are both criteria by which a single qubit and a single classical bit perform equivalently.

References

[1] A. Holevo: The Capacity of Quantum Channel with General Signal States. *IEEE Transactions on Information Theory* **44** (1998), pg. 269–273.