Quantum Information

Problem session 7
November 25th, 2022

Task 1: Holevo Bound

A concrete example involves Alice preparing a single qubit in one of two quantum states according to the outcome of a fair coin toss. If the coin toss yields heads, then Alice prepares the state $|0\rangle$, and if the coin toss yields tails, then Alice prepares the state $\cos\theta|0\rangle + \sin\theta|1\rangle$, where θ is some real parameter. In the $|0\rangle, |1\rangle$ basis it follows that ρ may be written

$$\rho = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}. \tag{12.15}$$

A simple calculation shows that the eigenvalues of ρ are $(1 \pm \cos\theta)/2$, and the Holevo bound is therefore given by the binary entropy $H((1 + \cos\theta)/2)$, as illustrated in Figure 12.1. Notice that the Holevo bound is maximized when $\theta = \pi/2$, attaining a value of 1 bit, corresponding to the case of Alice preparing states chosen from an orthogonal set, at which point it is possible for Bob to determine with surety which state Alice prepared. For other values of θ the Holevo bound is strictly less than 1 bit, and it is impossible for Bob to determine with surety which state Alice prepared.

Task 1: Holevo Bound

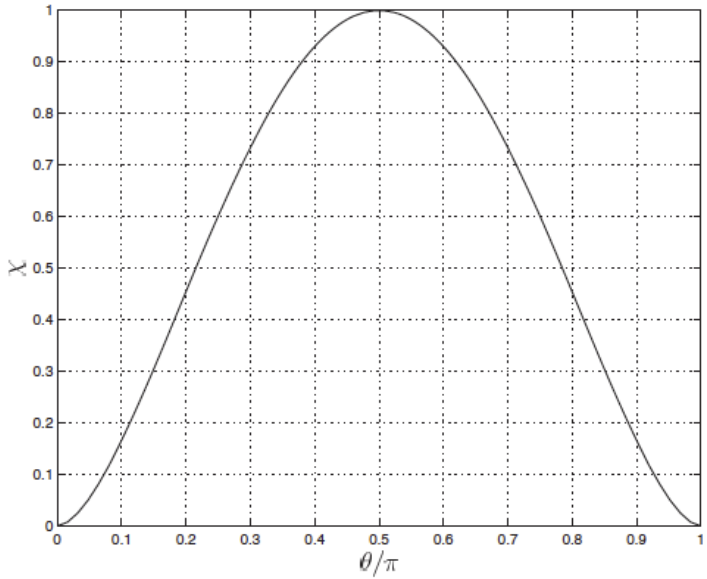

Figure 12.1. Plot of the Holevo bound χ as a function of θ when the states $|0\rangle$ and $\cos\theta|0\rangle + \sin\theta|1\rangle$ are prepared with equal probability. Notice that the Holevo bound reaches a maximum when $\theta = \pi/2$, corresponding to orthogonal states. It is only at this point that it is possible for Bob to determine with certainty which state Alice prepared.

Task 2: Example of Improved Distinguishability through Coding

To better acquaint ourselves with the concept of accessible information, let's consider a single-qubit example. Alice prepares one of the three possible pure states

$$|\varphi_1\rangle = |\uparrow_{\hat{n}_1}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$$|\varphi_2\rangle = |\uparrow_{\hat{n}_2}\rangle = \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix},$$

$$|\varphi_3\rangle = |\uparrow_{\hat{n}_3}\rangle = \begin{pmatrix} -\frac{1}{2} \\ -\frac{\sqrt{3}}{2} \end{pmatrix}; \tag{10.214}$$

a spin-$\frac{1}{2}$ object points in one of three directions that are symmetrically distributed in the xz-plane. Each state has *a priori* probability $\frac{1}{3}$. Evidently, Alice's signal states are nonorthogonal:

$$\langle\varphi_1|\varphi_2\rangle = \langle\varphi_1|\varphi_3\rangle = \langle\varphi_2|\varphi_3\rangle = -\frac{1}{2}. \tag{10.215}$$

Task 2: Improved Distinguishability through Coding

Bob's task is to find out as much as he can about what Alice prepared by making a suitable measurement. The density matrix of Alice's ensemble is

$$\boldsymbol{\rho} = \frac{1}{3}(|\varphi_1\rangle\langle\varphi_1| + |\varphi_2\rangle\langle\varphi_2| + |\varphi_3\rangle\langle\varphi_3|) = \frac{1}{2}\boldsymbol{I}, \tag{10.216}$$

which has $H(\boldsymbol{\rho}) = 1$. Therefore, the Holevo bound tells us that the mutual information of Alice's preparation and Bob's measurement outcome cannot exceed 1 bit.

In fact, though, the accessible information is considerably less than the one bit allowed by the Holevo bound. In this case, Alice's ensemble has enough symmetry that it is not hard to guess the optimal measurement. Bob may choose a POVM with three outcomes, where

$$\boldsymbol{E}_a = \frac{2}{3}(\boldsymbol{I} - |\varphi_a\rangle\langle\varphi_a|), \quad a = 1, 2, 3; \tag{10.217}$$

we see that

$$p(a|b) = \langle\varphi_b|\boldsymbol{E}_a|\varphi_b\rangle = \begin{cases} 0 & a = b, \\ \frac{1}{2} & a \neq b. \end{cases} \tag{10.218}$$

Task 2: Improved Distinguishability through Coding

The measurement outcome a *excludes* the possibility that Alice prepared a, but leaves equal *a posteriori* probabilities $\left(p = \frac{1}{2}\right)$ for the other two states. Bob's information gain is

$$I = H(X) - H(X|Y) = \log_2 3 - 1 = .58496. \tag{10.219}$$

To show that this measurement is really optimal, we may appeal to a variation on a theorem of Davies, which assures us that an optimal POVM can be chosen with three $\boldsymbol{E}_a$'s that share the same three-fold symmetry as the three states in the input ensemble. This result restricts the possible POVM's enough so that we can check that eq. (10.217) is optimal with an explicit calculation. Hence we have found that the ensemble $\mathcal{E} = \{|\varphi_a\rangle, p_a = \frac{1}{3}\}$ has accessible information.

$$\mathrm{Acc}(\mathcal{E}) = \log_2\left(\frac{3}{2}\right) = .58496... \tag{10.220}$$

The Holevo bound is not saturated.

Task 2: Improved Distinguishability through Coding

Now suppose that Alice has enough cash so that she can afford to send two qubits to Bob, where again each qubit is drawn from the ensemble $\mathcal{E}$. The obvious thing for Alice to do is prepare one of the *nine* states

$$|\varphi_a\rangle \otimes |\varphi_b\rangle, \quad a, b = 1, 2, 3, \tag{10.221}$$

each with $p_{ab} = 1/9$. Then Bob's best strategy is to perform the POVM eq. (10.217) on each of the two qubits, achieving a mutual information of .58496 bits per qubit, as before.

But, determined to do better, Alice and Bob decide on a different strategy. Alice will prepare one of *three* two-qubit states

$$|\Phi_a\rangle = |\varphi_a\rangle \otimes |\varphi_a\rangle, \quad a = 1, 2, 3, \tag{10.222}$$

each occurring with *a priori* probability $p_a = 1/3$. Considered one-qubit at a time, Alice's choice is governed by the ensemble $\mathcal{E}$, but now her two qubits have (classical) correlations – both are prepared the same way.

The three $|\Phi_a\rangle$'s are linearly independent, and so span a three-dimensional subspace

Task 2: Improved Distinguishability through Coding

of the four-dimensional two-qubit Hilbert space. In Exercise 10.4, you will show that the density operator

$$\rho = \frac{1}{3}\left(\sum_{a=1}^{3} |\Phi_a\rangle\langle\Phi_a|\right), \tag{10.223}$$

has the nonzero eigenvalues $1/2, 1/4, 1/4$, so that

$$H(\rho) = -\frac{1}{2}\log_2\frac{1}{2} - 2\left(\frac{1}{4}\log_2\frac{1}{4}\right) = \frac{3}{2}. \tag{10.224}$$

The Holevo bound requires that the accessible information *per qubit* is no more than 3/4 bit, which is at least consistent with the possibility that we can exceed the .58496 bits per qubit attained by the nine-state method.

Naively, it may seem that Alice won't be able to convey as much classical information to Bob, if she chooses to send one of only three possible states instead of nine. But on further reflection, this conclusion is not obvious. True, Alice has fewer signals to choose from, but the signals are *more distinguishable;* we have

$$\langle\Phi_a|\Phi_b\rangle = \frac{1}{4}, \quad a \neq b, \tag{10.225}$$

instead of eq. (10.215). It is up to Bob to exploit this improved distinguishability in his choice of measurement. In particular, Bob will find it advantageous to perform *collective* measurements on the two qubits instead of measuring them one at a time.

Task 2: Improved Distinguishability through Coding

It is no longer obvious what Bob's optimal measurement will be. But Bob can invoke a general procedure that, while not guaranteed optimal, is usually at least pretty good. We'll call the POVM constructed by this procedure a "pretty good measurement" (or PGM).

Consider some collection of vectors $|\tilde{\Phi}_a\rangle$ that are not assumed to be orthogonal or normalized. We want to devise a POVM that can distinguish these vectors reasonably well. Let us first construct

$$\boldsymbol{G} = \sum_a |\tilde{\Phi}_a\rangle\langle\tilde{\Phi}_a|; \tag{10.226}$$

This is a positive operator on the space spanned by the $|\tilde{\Phi}_a\rangle$'s. Therefore, on that subspace, $\boldsymbol{G}$ has an inverse, $\boldsymbol{G}^{-1}$ and that inverse has a positive square root $\boldsymbol{G}^{-1/2}$. Now we define

$$\boldsymbol{E}_a = \boldsymbol{G}^{-1/2}|\tilde{\Phi}_a\rangle\langle\tilde{\Phi}_a|\boldsymbol{G}^{-1/2}, \tag{10.227}$$

and we see that

$$\begin{aligned}\sum_a \boldsymbol{E}_a &= \boldsymbol{G}^{-1/2}\left(\sum_a |\tilde{\Phi}_a\rangle\langle\tilde{\Phi}_a|\right)\boldsymbol{G}^{-1/2} \\ &= \boldsymbol{G}^{-1/2}\boldsymbol{G}\boldsymbol{G}^{-1/2} = \boldsymbol{I},\end{aligned} \tag{10.228}$$

on the span of the $|\tilde{\Phi}_a\rangle$'s. If necessary, we can augment these $\boldsymbol{E}_a$'s with one more positive operator, the projection $\boldsymbol{E}_0$ onto the orthogonal complement of the span of the $|\tilde{\Phi}_a\rangle$'s, and so construct a POVM. This POVM is the PGM associated with the vectors $|\tilde{\Phi}_a\rangle$.

In the special case where the $|\tilde{\Phi}_a\rangle$'s are orthogonal,

$$|\tilde{\Phi}_a\rangle = \sqrt{\lambda_a}|\phi_a\rangle, \tag{10.229}$$

Task 2: Improved Distinguishability through Coding

(where the $|\phi_a\rangle$'s are orthonormal), we have

$$\begin{aligned} \boldsymbol{E}_a &= \sum_{b,c} (|\phi_b\rangle \lambda_b^{-1/2} \langle\phi_b|)(|\phi_a\rangle \lambda_a \langle\phi_a|)(|\phi_c\rangle \lambda_c^{-1/2} \langle\phi_c|) \\ &= |\phi_a\rangle\langle\phi_a|; \end{aligned} \tag{10.230}$$

this is the orthogonal measurement that perfectly distinguishes the $|\phi_a\rangle$'s and so clearly is optimal. If the $|\tilde{\Phi}_a\rangle$'s are linearly independent but not orthogonal, then the PGM is again an orthogonal measurement (because n one-dimensional operators in an n-dimensional space can constitute a POVM only if mutually orthogonal — see Exercise 3.11), but in that case the measurement may not be optimal.

In Exercise 10.4, you'll construct the PGM for the vectors $|\Phi_a\rangle$ in eq. (10.222), and you'll show that

$$\begin{aligned} p(a|a) &= \langle\Phi_a|\boldsymbol{E}_a|\Phi_a\rangle = \frac{1}{3}\left(1 + \frac{1}{\sqrt{2}}\right)^2 = .971405, \\ p(b|a) &= \langle\Phi_a|\boldsymbol{E}_b|\Phi_a\rangle = \frac{1}{6}\left(1 - \frac{1}{\sqrt{2}}\right)^2 = .0142977 \end{aligned} \tag{10.231}$$

(for $b \neq a$). It follows that the conditional entropy of the input is

$$H(X|Y) = .215894, \tag{10.232}$$

and since $H(X) = \log_2 3 = 1.58496$, the information gain is

$$I(X;Y) = H(X) - H(X|Y) = 1.369068, \tag{10.233}$$

a mutual information of .684534 bits per qubit. Thus, the improved distinguishability of Alice's signals has indeed paid off – we have exceeded the .58496 bits that can be extracted from a single qubit. We still didn't saturate the Holevo bound ($I \leq 1.5$ in this case), but we came a lot closer than before.

Task 2: Improved Distinguishability through Coding

This example, first described by Peres and Wootters, teaches some useful lessons.

First, Alice is able to convey more information to Bob by “pruning” her set of codewords. She is better off choosing among fewer signals that are more distinguishable than more signals that are less distinguishable. An alphabet of three letters encodes more than an alphabet of nine letters.

Second, Bob is able to read more of the information if he performs a collective measurement instead of measuring each qubit separately. His optimal orthogonal measurement projects Alice’s signal onto a basis of entangled states.