APPPHYS225 - Tuesday 30 September 2008

Quantum states and measurements as non-commutative probability

1. Observables as non-commutative random variables; algebras thereof
2. Spectral decomposition, projection operators as basis elements of the algebra
3. Density matrix as a representation of state
4. Commutative sub-algebras as classical probability models
5. Heisenberg Uncertainty Principle
6. Binary quantum state discrimination
7. Measures of distance/distinguishability for quantum states
8. Dirac notation review

Quantum states and measurements as non-commutative probability

Last week we saw that a classical probability model can be formulated in terms of an algebra of observables (random variables), and a state on that algebra. There is a natural representation of the observables as diagonal matrices, and the state can then be represented by a diagonal matrix ρ such that

$$\langle A \rangle = \mathrm{Tr}[(A)\rho].$$

The random variables can be viewed as functions on a sample space, with ρ then a representation of the probability distribution function, but this is not essential. We saw that conditioning can be performed directly on the algebra of observables (as opposed to the probability distribution function) via expressions such as

$$\langle E \rangle_F(\cdot) = \sum_j \frac{\langle E\chi_{f_j} \rangle}{\langle \chi_{f_j} \rangle} \chi_{f_j}(\cdot),$$

which can easily be mapped into the matrix representation. The indicator functions $\chi(\cdot)$ provide basis functions in the algebra of random variables, whose matrix representations are projection operators that similarly provide a basis for the matrix algebra of observables.

Today we take some (deceptively) simple steps to make the generalization to (discrete) quantum probability models. We retain the basic structure of an algebra of observables with a state, both of which will still have natural matrix representations, but we no longer require these to be diagonal. We will however require that quantum observables be *Hermitian*, in order to ensure that they take values in the real numbers only. Hence to begin with, we have the following simple table of equivalences:

□	***observables***	***state matrix*** ρ
classical models	diagonal matrices	diagonal, non-negative, $\mathrm{Tr} = 1$
quantum models	Hermitian matrices	Hermitian, non-negative, $\mathrm{Tr} = 1$

For diagonal matrices, non-negative simply means that the entries are ≥ 0. For

general Hermitian matrices, non-negative means that all eigenvalues are ≥ 0. In quantum probability models we retain the rule

$$\langle A \rangle = \mathrm{Tr}[A\rho],$$

where we have dropped the (A) notation as in the quantum case we will tend not to distinguish between an observable and its matrix representation. A famous example of a quantum probability model is that of a spin-$1/2$ particle, or *qubit*, which can be built from the following basis set of 2×2 observables:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Here σ_j is the observable corresponding to spin angular momentum along the j axis (we have taken $\hbar \to 1$). If we form the linear span of these over the real numbers we obtain Hermitian matrices of the form

$$M = \begin{pmatrix} a & c - id \\ c + id & b \end{pmatrix}, \quad a, b, c, d \in \mathbf{R},$$

which are observables contained within the algebra of 2×2 complex matrices. Our observables are easily seen to be non-commutative under multiplication since

$$\sigma_x\sigma_y - \sigma_y\sigma_x = 2i\sigma_z, \quad \sigma_y\sigma_z - \sigma_z\sigma_y = 2i\sigma_x, \quad \sigma_z\sigma_x - \sigma_x\sigma_z = 2i\sigma_y.$$

The set of valid state matrices for this model can be written

$$\rho = \frac{1}{2}(I + \langle\sigma_x\rangle\sigma_x + \langle\sigma_y\rangle\sigma_y + \langle\sigma_z\rangle\sigma_z) = \frac{1}{2}\begin{pmatrix} 1 + \langle\sigma_z\rangle & \langle\sigma_x\rangle - i\langle\sigma_y\rangle \\ \langle\sigma_x\rangle + i\langle\sigma_y\rangle & 1 - \langle\sigma_z\rangle \end{pmatrix},$$

where we are here thinking of $\langle\sigma_i\rangle$ as convenient parameters for the state matrix. We'll examine this model and its symmetries much further, later in the course.

It is worth noting that the set of valid quantum state matrices (like the set of valid classical state matrices) is closed under convex combination:

$$\rho = p\rho_1 + (1 - p)\rho_2,$$

where $0 \leq p \leq 1,$ or more generally

$$\rho = \sum_n p_n\rho_n, \quad 0 \leq p_n \leq 1, \qquad \sum_n p_n = 1.$$

Generally speaking, we can think of convex combination as a way of representing randomized preparation of the system. By adopting the state matrix $\rho = p\rho_1 + (1 - p)\rho_2$ we assign expectation values

$$\langle A \rangle = \mathrm{Tr}[A\rho] = p\mathrm{Tr}[A\rho_1] + (1 - p)\mathrm{Tr}[A\rho_2],$$

which we can think of as an average, weighted by p, over the values that would be obtained with ρ_1 and ρ_2. Some quantum states cannot be obtained as convex combinations of any others, and these are called pure states. They are rank-1 projectors. In classical probability the rank-1 projectors are state matrices corresponding to precise configurations, and there are only a limited number of such states for any finite sample space. In quantum probability there are a continuum of pure states for any matrix dimension.

For classical models we noted that the basis decomposition of a random variable in terms of indicator functions on its level sets,

$$A(\cdot) = \sum_i a_i \chi_{a_i}(\cdot),$$

takes the appearance in matrix representation of a spectral decomposition:

$$(A) = \sum_i a_i \Pi_{a_i},$$

where for example if $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and

$$A(\omega_1) = 2, \quad A(\omega_2) = A(\omega_3) = 3,$$

then

$$(A) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is clear that in the spectral decomposition the $\{a_i\}$ are the eigenvalues of (A) and the $\{\Pi_{a_i}\}$ are projectors onto the corresponding eigenspaces. In quantum models we use the fact that any Hermitian matrix has a similar spectral decomposition,

$$Q = \sum_i q_i \Pi_{q_i},$$

where the $\{q_i\}$ are the eigenvalues of Q and $\{\Pi_{q_i}\}$ are orthogonal projectors onto the corresponding eigenspaces. Recall that in the classical case we can compute the probability of $A = a_i$ as $\langle \chi_{a_i} \rangle = \mathrm{Tr}[\rho \Pi_{a_i}]$. In the quantum case we have the same probability rule, that in a measurement of Q we obtain the outcome q_i with probability $\langle \Pi_{q_i} \rangle = \mathrm{Tr}[\rho \Pi_{q_i}]$, which is consistent with the rule for expecations,

$$\langle Q \rangle = \mathrm{Tr}[\rho Q] = \mathrm{Tr}\left[\rho \sum_i q_i \Pi_{q_i} \right] = \sum_i q_i \mathrm{Tr}[\rho \Pi_{q_i}] = \sum_i q_i \Pr(q_i).$$

To illustrate some of these ideas let us consider the 3×3 example

$$Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

which is Hermitian and definitely not diagonal. What is its spectral decomposition? First we note that its eigenvalues and eigenvectors are

$$-1 : \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad 0 : \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad 1 : \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The projectors onto the eigenspaces are thus

$$\Pi_{-1} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

$$\Pi_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Pi_{+1} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Apparently then we are to interpret Q as an observable whose possible values are $\{-1, 0, 1\}$, and for which the non-diagonal projectors $\{\Pi_{-1}, \Pi_0, \Pi_{+1}\}$ generalize the role of indicator functions on level sets for classical random variables. Although the projectors are not diagonal they are still mutually orthogonal (they are guaranteed to be so since they correspond to distinct eigenvalues), so

$$\Pi_{-1}\Pi_0 = \Pi_{-1}\Pi_{+1} = \Pi_0\Pi_{+1} = 0.$$

It is worth noting that any Hermitian projection operator can be converted into a valid state matrix ρ for a quantum probability model via

$$\rho = \Pi/\mathrm{Tr}[\Pi].$$

In our current example all the eigenprojectors are rank-1 and we can easily see that $\rho = \Pi_\lambda$ is a state that assigns expectation value λ to Q :

$$\langle Q \rangle = \mathrm{Tr}[Q\rho] \to \mathrm{Tr}[Q\Pi_\lambda] = \mathrm{Tr}[(\Pi_{+1} - \Pi_{-1})\Pi_\lambda] = \lambda.$$

Hence we see that the existence of non-diagonal state matrices is somehow natural in a model containing observables “built from” non-diagonal projectors.

Of course, the statement that a matrix is non-diagonal actually depends on our choice of basis. Any Hermitian matrix is *diagonalizable*, meaning that we can express

$$Q = P\Lambda P^{-1},$$

where Λ is a diagonal matrix of eigenvalues and P is a matrix containing the corresponding eigenvectors as columns. In fact since Q is Hermitian it is unitarily diagonalizable, meaning that $P^{-1} = P^*$ (Hermitian conjugate) as long as we construct P using orthonormal eigenvectors. Recall that the eigenvectors of a Hermitian matrix can always be chosen to provide an orthonormal basis for the vector space the matrix acts on. For our example,

$$\Lambda = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad P = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad P^{-1} = P^* = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

$$P\Lambda P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

This means that if we consistently apply the similarity transformation $M \mapsto P^{-1}MP$

(which is actually just a unitary rotation $M \mapsto P^* M P$) to all matrices in our quantum probability model, we will actually obtain a diagonal representation for Q :

$$Q \mapsto P^* Q P = P^* (P \Lambda P^*) P = \Lambda.$$

At the same time, we must transform $\rho \mapsto P^* \rho P$ and likewise for any other observable we are concerned with. Note that any power of Q also becomes diagonal with this change of basis:

$$Q^n = (P \Lambda P^*)^n = P \Lambda^n P^* \mapsto \Lambda^n.$$

The set of all powers of Q actually provides a basis for a closed algebra of observables:

$$\alpha \left(\sum_{m=0}^{\infty} c_m Q^m \right) + \beta \left(\sum_{n=0}^{\infty} d_n Q^n \right) = \sum_{p=0}^{\infty} (\alpha c_p + \beta d_p) Q^p, \quad \alpha, \beta \in \mathbf{R},$$

$$\gamma \left(\sum_{m=0}^{\infty} c_m Q^m \right) \left(\sum_{n=0}^{\infty} d_n Q^n \right) = \sum_{p=0}^{\infty} \gamma \left(\sum_{\{m,n \,:\, m+n=p\}} c_m d_n \right) Q^p, \quad \gamma \in \mathbf{R},$$

all of which are diagonal in the eigenbasis of Q :

$$\sum_{p=0}^{\infty} c_p Q^p \mapsto \sum_{p=0}^{\infty} c_p \Lambda^p.$$

We thus find that if we are only interested in observables in this closed diagonal algebra, it appears that we can choose a basis that makes them look equivalent to observables of a classical probability model. And what about the state matrix? Let us first note

$$\tilde{\Pi}_{-1} \equiv P^* \Pi_{-1} P = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\tilde{\Pi}_0 \equiv P^* \Pi_0 P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \tilde{\Pi}_{+1} \equiv P^* \Pi_{+1} P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and introduce the notation

$$\tilde{\rho} \equiv P^* \rho P.$$

Then as long as we are only concerned with computing expectation values for observables in our restricted algebra, we see that

$$\left\langle \sum_{p=0}^{\infty} c_p \Lambda^p \right\rangle = \sum_{p=0}^{\infty} c_p \langle \Lambda^p \rangle = \sum_{p=0}^{\infty} c_p \mathrm{Tr}[\Lambda^p \tilde{\rho}] = \sum_{p=0}^{\infty} c_p \left((-1)^p \mathrm{Tr}\left[\tilde{\Pi}_{-1} \tilde{\rho} \right] + 0^p \mathrm{Tr}\left[\tilde{\Pi}_0 \tilde{\rho} \right] + 1^p \mathrm{Tr}\left[\tilde{\Pi}_{+1} \tilde{\rho} \right] \right)$$

$$= \sum_{p=0}^{\infty} c_p \left((-1)^p \mathrm{Tr}\left[\tilde{\Pi}_{-1} \tilde{\rho} \tilde{\Pi}_{-1} \right] + 0^p \mathrm{Tr}\left[\tilde{\Pi}_0 \tilde{\rho} \tilde{\Pi}_0 \right] + 1^p \mathrm{Tr}\left[\tilde{\Pi}_{+1} \tilde{\rho} \tilde{\Pi}_{+1} \right] \right)$$

$$= \sum_{p=0}^{\infty} c_p ((-1)^p (\tilde{\rho})_{11} + 0^p (\tilde{\rho})_{22} + 1^p (\tilde{\rho})_{33}),$$

where $(\tilde{\rho})_{ij}$ denotes the (i,j) matrix element of $\tilde{\rho}$. Hence we see that the expectation values of observables in the restricted algebra, and hence the *state* on the restricted algebra, depends only on the diagonal elements of the transformed state matrix.

Clearly this argument about the existence of an equivalent classical probability model generalizes to any algebra of observables that can be simulataneously diagonalized. As we know from basic linear algebra this means that any *commutative* algebra of observables, even if we want to think of them as being "quantum" in origin, has an equivalent classical probability model. If among the set of observables we are concerned with there exist any pairs that do not commute, however, then there does not exist a simultaneously-diagonalizing linear transformation. Hence we see the existence of non-commuting observables as fundamental to the distinction between classical and quantum probability.

One often hears that the primary significance of the existence of non-commuting observables is the Heisenberg Uncertainty Principle. Here we take a moment to ponder this assertion. Consider a pair of observables A, B that do not commute, and let

$$[A,B] = AB - BA \equiv iC,$$

where $A,\ B$ and C are all Hermitian. Now consider an arbitrary state matrix ρ and define

$$\tilde{A} \equiv A - \langle A \rangle, \quad \tilde{B} \equiv B - \langle B \rangle,$$

so that

$$\langle \tilde{A}^2 \rangle = \left\langle (A - \langle A \rangle)^2 \right\rangle = \langle A^2 \rangle - \langle A \rangle^2 = (\Delta A)^2,$$

$$\langle \tilde{B}^2 \rangle = \left\langle (B - \langle B \rangle)^2 \right\rangle = \langle B^2 \rangle - \langle B \rangle^2 = (\Delta B)^2,$$

$$\langle \tilde{A}\tilde{B} \rangle = \langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle = \langle AB \rangle - \langle A \rangle \langle B \rangle.$$

Note that $(A,B) \equiv \mathrm{Tr}[\rho A^* B]$ is an inner product and therefore satisfies a Cauchy-Schwartz inequality:

$$|(A,B)|^2 \leq (A,A)(B,B),$$

which in our scenario with Hermitian matrices implies

$$\left| \mathrm{Tr}[\rho \tilde{A}\tilde{B}] \right|^2 \leq \left| \mathrm{Tr}[\rho \tilde{A}^2] \right| \left| \mathrm{Tr}[\rho \tilde{B}^2] \right|,$$

$$|\langle AB \rangle - \langle A \rangle \langle B \rangle| \leq \Delta A \Delta B.$$

Now we massage this a bit further, with

$$\begin{aligned} AB &= \frac{1}{2}[(AB+BA)+(AB-BA)] \\ &= \frac{1}{2}[(AB+BA)+iC], \\ \langle AB\rangle &= \frac{1}{2}[\langle AB+BA\rangle + i\langle C\rangle]. \end{aligned}$$

Since $AB+BA$ and C are both Hermitian operators, $\langle AB+BA\rangle$ is real while $i\langle C\rangle$ is purely imaginary. Hence

$$|\langle AB\rangle - \langle A\rangle\langle B\rangle| = \sqrt{\left(\frac{1}{2}\langle AB+BA\rangle - \langle A\rangle\langle B\rangle\right)^2 + \frac{1}{4}\langle C\rangle^2} \geq \frac{1}{2}|\langle C\rangle|,$$

and thus

$$\Delta A \Delta B \geq \frac{1}{2}|\langle [A,B]\rangle|.$$

which is the usual statement of the Heisenberg Uncertainty Principle. This is of course often a useful quantitative expression, but does it say anything qualitatively profound about the consequences of non-commutativity?

The first thing to notice is that there are always states that make the RHS of the inequality vanish. If we write the spectral decomposition for A,

$$A = \sum_i a_i \Pi_i,$$

and recall that the $\{\Pi_i\}$ are mutually orthogonal, we can pick any eigenprojector of A as state matrix:

$$\rho_i = \Pi_i / \mathrm{Tr}[\Pi_i].$$

With this choice of state,

$$\begin{aligned} \langle [A,B]\rangle &= \mathrm{Tr}[AB\rho_i] - \mathrm{Tr}[BA\rho_i] \\ &= \mathrm{Tr}[B\rho_i A] - \mathrm{Tr}[B\rho_i A] \\ &= 0, \end{aligned}$$

where we have used both the cyclic property of Tr and the fact that

$$\Pi_i A = \Pi_i \left(\sum_j a_j \Pi_j\right) = A\Pi_i.$$

Hence *in finite dimensions* there is no non-trivial global lower bound on the uncertainty product of two observables, even if they do not commute. This makes sense of course since any eigenstate of A gives us $\Delta A = 0,$ so as long as ΔB is bounded we would expect to have vanishing uncertainty product (and similarly for an eigenstate of B). Now in infinite dimensions we have for example

$$[x,p] = i\hbar,$$

which means that there are no states for which the uncertainty product $\Delta x \Delta p$ vanishes - apparently we should think of this as related to the unboundedness of operators such as x and p, or perhaps to the fact that the eigenstates of x and p aren't really valid quantum states (as they aren't square-normalizable).

In a classical probability model we can find combinations of states and

observables that have non-zero uncertainty product. For example, if we take

$$C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \rho = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix},$$

we have

$$\Delta C_1 = \sqrt{\langle C_1^2 \rangle - \langle C_1 \rangle^2} = \sqrt{\frac{1}{2} - \frac{1}{4}} = \frac{1}{2},$$

$$\Delta C_2 = \sqrt{\langle C_2^2 \rangle - \langle C_2 \rangle^2} = \sqrt{\frac{1}{2} - \frac{1}{4}} = \frac{1}{2},$$

so $\Delta C_1 \Delta C_2 = 1/4$ for this choice of ρ. Hence we confirm that the existence of non-zero uncertainty products is not a unique aspect of quantum probability.

Note however that in the classical setting we are forced to use non-pure states as soon as we ask for non-zero uncertainty even of a single observable. In the quantum setting on the other hand,

$$\rho = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad Q^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\Delta Q = \sqrt{\langle Q^2 \rangle - \langle Q \rangle^2} = \sqrt{1 - 0} = 1.$$

Hence we see that one key consequence of non-commutativity (which, as we saw earlier, requires us to consider non-diagonal states/observables) is that there exist pure states in the model that have non-zero uncertainty for some observables. This type of pure-state uncertainty in the value of an observable is often referred to as intrinsic quantum uncertainty, as we cannot view it as being the result of a randomized preparation.

Last week we reviewed classical probability formalism for handling joint systems and conditioning; we will defer a parallel discussion for quantum probability models until a bit later in the term.

Binary quantum state discrimination

Suppose I prepare a quantum system either in the state ρ_1 or the state ρ_2, with even probability, but I don't tell you which. I give you the physical system and allow you to make a single measurement, on the basis of which you must try to guess whether I prepared ρ_1 or ρ_2. This is the quantum analogue of the classical state discrimination problem we considered last week using conditional probability. Now, we haven't yet discussed conditioning for quantum states, so you might worry that we wouldn't be ready yet to take on the quantum state discrimination problem. But in fact we already know how to do everything we need to do!

For example suppose we are working in the qubit model with states

$$\rho_1 = \begin{pmatrix} a_1 & c_1 - id_1 \\ c_1 + id_1 & 1 - a_1 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} a_2 & c_2 - id_2 \\ c_2 + id_2 & 1 - a_2 \end{pmatrix},$$

and assume we have chosen the labels such that $a_1 > a_2$. Suppose that I choose between these by flipping a fair coin, let $M(\bullet)$ be a random variable on the sample space of the coin $\Omega = \{\omega_1, \omega_2\}$ such that $M(\omega_1) = 1,\ M(\omega_2) = 2$, and suppose that I prepare the quantum system in state ρ_m according the value of $M = m$. Note that before you make any measurements, your state on the algebra of observables is

$$\rho = \frac{1}{2}(\rho_1 + \rho_2),$$

since I have promised even probabilities for the two different preparations. If you choose for example the observable

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

the possible outcomes and associated projection operators are

$$+1 \leftrightarrow \Pi_{+z} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad -1 \leftrightarrow \Pi_{-z} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The set of forward probabilities are thus

$$\Pr(\sigma_z = +1 | m = 1) = \mathrm{Tr}[\rho_1 \Pi_{+z}] = a_1, \quad \Pr(\sigma_z = +1 | m = 2) = \mathrm{Tr}[\rho_2 \Pi_{+z}] = a_2,$$
$$\Pr(\sigma_z = -1 | m = 1) = 1 - a_1, \quad \Pr(\sigma_z = -1 | m = 2) = 1 - a_2.$$

The conditional probabilites are then

$$\Pr(m = 1 | \sigma_z = +1) = \frac{\Pr(\sigma_z = +1 | m = 1)\Pr(m = 1)}{\Pr(\sigma_z = +1)} = \frac{a_1(\frac{1}{2})}{\frac{1}{2}(a_1 + a_2)} = \frac{a_1}{a_1 + a_2},$$
$$\Pr(m = 2 | \sigma_z = +1) = \frac{a_2}{a_1 + a_2},$$
$$\Pr(m = 1 | \sigma_z = -1) = \frac{1 - a_1}{2 - a_1 - a_2},$$
$$\Pr(m = 2 | \sigma_z = -1) = \frac{1 - a_2}{2 - a_1 - a_2}.$$

Hence assuming $a_1 > a_2$, we should guess ρ_1 if the result is $\sigma_z = +1$ and ρ_2 if the result is $\sigma_z = -1$. The probability of error is then

$$\mathrm{PE}_z = \Pr(\sigma_z = -1 | m = 1)\Pr(m = 1) + \Pr(\sigma_z = +1 | m = 2)\Pr(m = 2) = \frac{1}{2}(1 - a_1 + a_2).$$

As we might expect, the only way for this to vanish is if $a_1 = 1$ and $a_2 = 0$.

Suppose however that you had chosen to measure σ_x rather than σ_z. In that case we would utilize the spectral decomposition

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \Pi_{+x} - \Pi_{-x} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

and obtain the forward probabilities

$$\Pr(\sigma_x = +1|m) = \mathrm{Tr}[\rho_m \Pi_{+x}] = \frac{1}{2}\mathrm{Tr}\left[\begin{pmatrix} a_m & c_m - id_m \\ c_m + id_m & 1 - a_m \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right] = \frac{1}{2} + c_m,$$

$$\Pr(\sigma_x = -1|m) = \frac{1}{2} - c_m,$$

and the conditional probabilities

$$\Pr(m|\sigma_x = +1) = \frac{\Pr(\sigma_x = +1|m)\Pr(m)}{\Pr(\sigma_x = +1)} = \frac{(\frac{1}{2} + c_m)\frac{1}{2}}{\frac{1}{2}(1 + c_1 + c_2)} = \frac{\frac{1}{2} + c_m}{1 + c_1 + c_2},$$

$$\Pr(m|\sigma_x = -1) = \frac{\frac{1}{2} - c_m}{1 - c_1 - c_2}.$$

Let us suppose, just for the sake of simplifying the argument, that $c_1 > c_2$. Then we should guess ρ_1 if the outcome is $\sigma_x = +1$ and ρ_2 if the outcome is $\sigma_x = -1$, and the probability of error is

$$\mathrm{PE}_x = \Pr(\sigma_x = -1|m = 1)\Pr(m = 1) + \Pr(\sigma_x = +1|m = 2)\Pr(m = 2) = \frac{1}{2}(1 - c_1 + c_2).$$

Hence a measurement of σ_x is better than a measurement of σ_z if

$$\mathrm{PE}_x < \mathrm{PE}_z,$$

$$1 - c_1 + c_2 < 1 - a_1 + a_2,$$

$$c_2 - c_1 < a_2 - a_1.$$

Since we are assuming $a_1 > a_2$ and $c_1 > c_2$, and recalling

$$a_1 = \frac{1}{2}(1 + \langle\sigma_z\rangle_1), \quad a_2 = \frac{1}{2}(1 + \langle\sigma_z\rangle_2), \quad c_1 = \frac{1}{2}\langle\sigma_x\rangle_1, \quad c_2 = \frac{1}{2}\langle\sigma_x\rangle_2,$$

we see that $\mathrm{PE}_x < \mathrm{PE}_z$ if

$$\langle\sigma_x\rangle_1 - \langle\sigma_x\rangle_2 > \langle\sigma_z\rangle_1 - \langle\sigma_z\rangle_2.$$

Hence we should pick the observable for which the difference in expectation values between ρ_1 and ρ_2 is greater. Generally speaking, we expect that in any quantum state discrimination scenario it will be important to pick an optimal observable.

In any case we see that, just as in the classical case, some pairs of quantum states are easier to discriminate than others. As discussed in [C. A. Fuchs and J. van de Graaf, IEEE Transactions on Information Theory, Vol. 45, p. 1216 (1999)], measures of distinguishability can be derived for quantum states as well as classical:

$$\mathrm{PE}(\rho_1, \rho_2) = \frac{1}{2} - \frac{1}{4}\mathrm{Tr}|\rho_1 - \rho_2|,$$

$$\mathrm{K}(\rho_1, \rho_2) = \frac{1}{2}\mathrm{Tr}|\rho_1 - \rho_2|,$$

$$\mathrm{B}(\rho_1, \rho_2) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho_1}\,\rho_2\sqrt{\rho_1}}\,\right].$$

In the Probability of Error and Kolmogorov distance definitions, it is remarkably possible to write expressions directly in terms of the state matrices that already take into account an optimization over all possible measurements that could be used for the state discrimination. Next time we'll see something about how that kind of optimization can be performed.

Review of Dirac notation

Having made our case for the formal similarities between classical and quantum probability, we will now transition into more traditional quantum-mechanics notation for our upcoming discussions of subtleties. Most of you are probably quite familiar with Dirac notation, but here we'll include a brief review just to make sure everyone is on the same page.

So far we have represented quantum observables and states as $N \times N$ Hermitian matrices. Any such matrix implicitly acts upon an N-dimensional complex vector space H_N. Vectors in this vector space are denoted by Dirac kets and bras:

$$|\Psi_c\rangle \leftrightarrow \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}, \quad |\Psi_c\rangle^* = \langle\Psi_c| = \begin{pmatrix} c_1^* & c_2^* & \cdots & c_N^* \end{pmatrix}.$$

Inner (scalar) products can then be denoted as follows:

$$\langle\Psi_c|\Psi_d\rangle \equiv \langle\Psi_c||\Psi_d\rangle = \begin{pmatrix} c_1^* & c_2^* & \cdots & c_N^* \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix} = \sum_{i=1}^{N} c_i^* d_i,$$

with $\langle\Psi_c|\Psi_c\rangle = 1$ indicating that $|\Psi_c\rangle$ is normalized and $\langle\Psi_c|\Psi_d\rangle = 0$ indicating that $|\Psi_c\rangle$ and $|\Psi_d\rangle$ are orthogonal. It is natural to choose an orthonormal basis for the vector space

$$H_N = \text{span}\{|1\rangle, |2\rangle, \ldots, |N\rangle\}, \quad \langle i|j\rangle = \delta_{ij},$$

in terms of which we can write expansions

$$|\Psi_c\rangle = \sum_i c_i |i\rangle, \quad \langle\Psi_c| = \sum_i c_i^* \langle i|.$$

The trace of a matrix O can likewise be written

$$\text{Tr}[O] = \sum_i \langle i|O|i\rangle.$$

Outer products can be defined via

$$|\Psi_c\rangle\langle\Psi_d| = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix} \begin{pmatrix} d_1^* & d_2^* & \cdots & d_N^* \end{pmatrix} = \begin{pmatrix} c_1 d_1^* & c_2 d_1^* & \cdots & c_N d_1^* \\ c_1 d_2^* & c_2 d_2^* & & c_N d_2^* \\ \vdots & & \ddots & \vdots \\ c_1 d_N^* & c_2 d_N^* & \cdots & c_N d_N^* \end{pmatrix}.$$

We note that if $|\Psi_c\rangle$ represents a normalized vector, then

$$|\Psi_c\rangle\langle\Psi_c| = \begin{pmatrix} |c_1|^2 & c_2c_1^* & \cdots & c_Nc_1^* \\ c_1c_2^* & |c_2|^2 & & c_Nc_2^* \\ \vdots & & \ddots & \vdots \\ c_1c_N^* & c_2c_N^* & \cdots & |c_N|^2 \end{pmatrix}$$

is a rank-1 orthogonal projector onto the axis defined by $|\Psi_c\rangle$. We can see that it is a projector, for example, by computing

$$(|\Psi_c\rangle\langle\Psi_c|)^2 = |\Psi_c\rangle(\langle\Psi_c\|\Psi_c\rangle)\langle\Psi_c| = |\Psi_c\rangle\langle\Psi_c|.$$

We can construct higher-rank projectors onto hyperplanes, etc., by combining orthogonal rank-1 projectors:

$$\Pi_{cd} = |\Psi_c\rangle\langle\Psi_c| + |\Psi_d\rangle\langle\Psi_d|, \quad \langle\Psi_c|\Psi_d\rangle = 0.$$

Again we can check,

$$\Pi_{cd}^2 = (|\Psi_c\rangle\langle\Psi_c| + |\Psi_d\rangle\langle\Psi_d|)(|\Psi_c\rangle\langle\Psi_c| + |\Psi_d\rangle\langle\Psi_d|) = \Pi_{cd},$$

and the need to have $\langle\Psi_c|\Psi_d\rangle = 0$ is clear.

The state matrix for a pure state can thus be written

$$\rho = |\Psi_c\rangle\langle\Psi_c|,$$

and it is common to think of the ket $|\Psi_c\rangle$ or the bra $\langle\Psi_c|$ as a natural vector representation of the state. For any given operator/matrix O,

$$O = \begin{pmatrix} o_{11} & o_{12} & \cdots & o_{1N} \\ o_{21} & o_{22} & & o_{2N} \\ \vdots & & \ddots & \vdots \\ o_{N1} & o_{N2} & \cdots & o_{NN} \end{pmatrix} = \sum_{i,j} o_{ij}|i\rangle\langle j|,$$

we have

$$\langle O\rangle = \mathrm{Tr}[O\rho] = \mathrm{Tr}\left[\sum_{i,j} o_{ij}|i\rangle\langle j|\Psi_c\rangle\langle\Psi_c|\right] = \sum_k \langle k|\left(\sum_{i,j} o_{ij}|i\rangle\langle j|\Psi_c\rangle\langle\Psi_c|\right)|k\rangle$$

$$= \sum_{i,j,k} o_{ij}c_k^*c_j\delta_{k,i} = \sum_{i,j} o_{ij}c_i^*c_j = \langle\Psi_c|O|\Psi_c\rangle.$$

Finally we note that eigenvalue equations for Hermitian observables take the following form in Dirac notation:

$$T|\theta_i\rangle = t_i|\theta_{i,j}\rangle, \quad \langle\theta_{i,j}|\theta_{k,l}\rangle = \delta_{ik}\delta_{jl},$$

where the notation accounts for the fact that eigenvalues can be degenerate, and we assume that the eigenvectors have been orthonormalized. We then have the spectral decomposition

$$T = \sum_{i,j} t_i|\theta_{i,j}\rangle\langle\theta_{i,j}|.$$

If any of the eigenvalues t_i is degenerate, we can obviously group together the corresponding rank-1 projectors onto eigenvectors into a higher-rank projector onto

the entire eigenspace:

$$T = \sum_i t_i \Pi_i, \quad \Pi_i = \sum_j |\theta_{i,j}\rangle\langle\theta_{i,j}|.$$