Classical information

Like we did in the previous lesson, we'll begin this lesson with a discussion of classical information. Once again, the probabilistic and quantum descriptions are mathematically similar, and recognizing how the mathematics works in the familiar setting of classical information is helpful in understanding why quantum information is described in the way that it is.

Classical states via the Cartesian product

We'll start at a very basic level, with classical states of multiple systems. For simplicity, we'll begin by discussing just two systems, and then generalize to more than two systems.

To be precise, let X be a system whose classical state set is Σ, and let Y be a second system whose classical state set is Γ. Note that, because we have referred to these sets as *classical state sets*, our assumption is that Σ and Γ are both finite and nonempty. It could be that $\Sigma = \Gamma$, but this is not necessarily so — and regardless, it will be helpful to use different names to refer to these sets in the interest of clarity.

Now imagine that the two systems, X and Y, are placed side-by-side, with X on the left and Y on the right. If we so choose, we can view these two systems as if they form a single system, which we can denote by (X, Y) or XY depending on our preference. A natural question to ask about this compound system (X, Y) is, "What are its classical states?"

The answer is that the set of classical states of (X, Y) is the *Cartesian product* of Σ and Γ, which is the set defined as

$$\Sigma \times \Gamma = \big\{(a, b) \,:\, a \in \Sigma \text{ and } b \in \Gamma\big\}.$$

In simple terms, the Cartesian product is precisely the mathematical notion that captures the idea of viewing an element of one set and an

element of a second set together, as if they form a single element of a single set. In the case at hand, to say that (X, Y) is in the classical state $(a, b) \in \Sigma \times \Gamma$ means that X is in the classical state $a \in \Sigma$ and Y is in the classical state $b \in \Gamma$; and if the classical state of X is $a \in \Sigma$ and the classical state of Y is $b \in \Gamma$, then the classical state of the joint system (X, Y) is (a, b).

For more than two systems, the situation generalizes in a natural way. If we suppose that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are systems having classical state sets $\Sigma_1, \ldots, \Sigma_n$, respectively, for any positive integer n, the classical state set of the n-tuple $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, viewed as a single joint system, is the Cartesian product

$$\Sigma_1 \times \cdots \times \Sigma_n = \bigl\{(a_1, \ldots, a_n) \,:\, a_1 \in \Sigma_1,\ \ldots,\ a_n \in \Sigma_n\bigr\}.$$

Of course, we are free to use whatever names we wish for systems, and to order them as we choose. In particular, if we have n systems like above, we could instead choose to name them $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ and arrange them from right to left, so that the joint system becomes $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$. Following the same pattern for naming the associated classical states and classical state sets, we might then refer to a classical state

$$(a_{n-1}, \ldots, a_0) \in \Sigma_{n-1} \times \cdots \times \Sigma_0$$

of this compound system. Indeed, this is the ordering convention used by Qiskit when naming multiple qubits. We'll come back to this convention and how it connects to quantum circuits in the next lesson, but we'll start using it now to help to get used to it.

It is often convenient to write a classical state of the form $(a_{n-1}, \ldots, a_0)$ as a string $a_{n-1} \cdots a_0$ for the sake of brevity, particularly in the very typical situation that the classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$ are associated with sets of *symbols* or *characters*. In this context, the term *alphabet* is commonly used to refer to sets of symbols used to form strings, but the mathematical definition of an alphabet is precisely the same as the definition of a classical state set: it is a finite and nonempty set.

For example, suppose that $\mathsf{X}_0, \ldots, \mathsf{X}_9$ are bits, so that the classical state sets of these systems are all the same.

$$\Sigma_0 = \Sigma_1 = \cdots = \Sigma_9 = \{0, 1\}$$

There are then $2^{10} = 1024$ classical states of the joint system $(\mathsf{X}_9, \ldots, \mathsf{X}_0)$, which are the elements of the set

$$\Sigma_9 \times \Sigma_8 \times \cdots \times \Sigma_0 = \{0,1\}^{10}.$$

Written as strings, these classical states look like this:

$$\begin{array}{c} 0000000000 \\ 0000000001 \\ 0000000010 \\ 0000000011 \\ 0000000100 \\ \vdots \\ 1111111111 \end{array}$$

For the classical state 0000000110, for instance, we see that X_1 and X_2 are in the state 1, while all other systems are in the state 0.

Probabilistic states

Recall from the previous lesson that a *probabilistic state* associates a probability with each classical state of a system. Thus, a probabilistic state of multiple systems — viewed collectively as a single system — associates a probability with each element of the Cartesian product of the classical state sets of the individual systems.

For example, suppose that X and Y are both bits, so that their corresponding classical state sets are $\Sigma = \{0,1\}$ and $\Gamma = \{0,1\}$, respectively. Here is a probabilistic state of the pair (X,Y):

$$\begin{aligned} \mathrm{Pr}\big((\mathsf{X},\mathsf{Y}) = (0,0)\big) &= 1/2 \\ \mathrm{Pr}\big((\mathsf{X},\mathsf{Y}) = (0,1)\big) &= 0 \\ \mathrm{Pr}\big((\mathsf{X},\mathsf{Y}) = (1,0)\big) &= 0 \\ \mathrm{Pr}\big((\mathsf{X},\mathsf{Y}) = (1,1)\big) &= 1/2 \end{aligned}$$

This probabilistic state is one in which both X and Y are random bits — each is 0 with probability $1/2$ and 1 with probability $1/2$ — but the classical states of the two bits always agree. This is an example of a *correlation* between these systems.

Ordering Cartesian product state sets

Probabilistic states of systems can be represented by probability vectors, as was discussed in the previous lesson. In particular, the vector entries represent probabilities for the system to be in the possible classical states of that system, and the understanding is that a correspondence between the entries and the set of classical states has been selected.

Choosing such a correspondence effectively means deciding on an ordering of the classical states, which is often natural or determined by a standard convention. For example, the binary alphabet $\{0, 1\}$ is naturally ordered with 0 first and 1 second, so the first entry in a probability vector representing a probabilistic state of a bit is the probability for it to be in the state 0, and the second entry is the probability for it to be in the state 1.

None of this changes in the context of multiple systems, but there is a decision to be made. The classical state set of multiple systems together, viewed collectively as a single system, is the Cartesian product of the classical state sets of the individual systems — so we must decide how the elements of Cartesian products of classical state sets are to be ordered.

There is a simple convention that we follow for doing this, which is to start with whatever orderings are already in place for the individual classical state sets, and then to order the elements of the Cartesian product *alphabetically.* Another way to say this is that the entries in each n-tuple (or, equivalently, the symbols in each string) are treated as though they have significance that *decreases from left to right.* For example, according to this convention, the Cartesian product $\{1, 2, 3\} \times \{0, 1\}$ is ordered like this:

$$(1, 0),\ (1, 1),\ (2, 0),\ (2, 1),\ (3, 0),\ (3, 1).$$

When n-tuples are written as strings and ordered in this way, we observe familiar patterns, such as $\{0, 1\} \times \{0, 1\}$ being ordered as $00, 01, 10, 11$, and the set $\{0, 1\}^{10}$ being ordered as it was written earlier in the lesson. As another example, viewing the set $\{0, 1, \ldots, 9\} \times \{0, 1, \ldots, 9\}$ as a set of strings, we obtain the two-digit numbers 00 through 99, ordered numerically. This is obviously not a coincidence; our decimal number system uses precisely this sort of alphabetical ordering, where the word *alphabetical* should be understood as having a broad meaning that includes numerals in addition to letters.

Returning to the example of two bits from above, the probabilistic state described previously is therefore represented by the following probability vector, where the entries are labeled explicitly for the sake of clarity.

$$\begin{pmatrix} \frac{1}{2} \\ 0 \\ 0 \\ \frac{1}{2} \end{pmatrix} \begin{array}{l} \leftarrow \text{probability of being in the state } 00 \\ \leftarrow \text{probability of being in the state } 01 \\ \leftarrow \text{probability of being in the state } 10 \\ \leftarrow \text{probability of being in the state } 11 \end{array} \tag{1}$$

Independence of two systems

A special type of probabilistic state of two systems is one in which the systems are *independent*. Intuitively speaking, two systems are independent if learning the classical state of either system has no effect on the probabilities associated with the other. That is, learning what classical state one of the systems is in provides no information at all about the classical state of the other.

To define this notion precisely, let us suppose once again that X and Y are systems having classical state sets Σ and Γ, respectively. With respect to a given probabilistic state of these systems, they are said to be *independent* if it is the case that

$$\operatorname{Pr}((\mathsf{X},\mathsf{Y}) = (a,b)) = \operatorname{Pr}(\mathsf{X} = a)\operatorname{Pr}(\mathsf{Y} = b) \tag{2}$$

for every choice of $a \in \Sigma$ and $b \in \Gamma$.

To express this condition in terms of probability vectors, assume that the given probabilistic state of (X,Y) is described by a probability vector, written in the Dirac notation as

$$\sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle.$$

The condition (2) for independence is then equivalent to the existence of two probability vectors

$$|\phi\rangle = \sum_{a\in\Sigma} q_a|a\rangle \quad \text{and} \quad |\psi\rangle = \sum_{b\in\Gamma} r_b|b\rangle, \tag{3}$$

representing the probabilities associated with the classical states of X and Y, respectively, such that

$$p_{ab} = q_a r_b \tag{4}$$

for all $a \in \Sigma$ and $b \in \Gamma$.

For example, the probabilistic state of a pair of bits (X,Y) represented by the vector

$$\frac{1}{6}|00\rangle + \frac{1}{12}|01\rangle + \frac{1}{2}|10\rangle + \frac{1}{4}|11\rangle$$

is one in which X and Y are independent. Specifically, the condition required for independence is true for the probability vectors

$$|\phi\rangle = \frac{1}{4}|0\rangle + \frac{3}{4}|1\rangle \quad \text{and} \quad |\psi\rangle = \frac{2}{3}|0\rangle + \frac{1}{3}|1\rangle.$$

For instance, to make the probabilities for the 00 state match, we need $\frac{1}{6} = \frac{1}{4} \times \frac{2}{3}$, and indeed this is the case. Other entries can be verified in a similar manner.

On the other hand, the probabilistic state (1), which we may write as

$$\frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle, \tag{5}$$

does not represent independence between the systems X and Y. A simple way to argue this follows.

Suppose that there did exist probability vectors $|\phi\rangle$ and $|\psi\rangle$, as in equation (3) above, for which the condition (4) is satisfied for every choice of a and b. It would then necessarily be that

$$q_0 r_1 = \Pr\big((\mathsf{X}, \mathsf{Y}) = (0, 1)\big) = 0.$$

This implies that either $q_0 = 0$ or $r_1 = 0$, because if both were nonzero, the product $q_0 r_1$ would also be nonzero. This leads to the conclusion that either $q_0 r_0 = 0$ (in case $q_0 = 0$) or $q_1 r_1 = 0$ (in case $r_1 = 0$). We see, however, that neither of those equalities can be true because we must have $q_0 r_0 = 1/2$ and $q_1 r_1 = 1/2$. Hence, there do not exist vectors $|\phi\rangle$ and $|\psi\rangle$ satisfying the property required for independence.

Having defined independence between two systems, we can now define what is meant by *correlation:* it is a *lack of independence*. For example, because the two bits in the probabilistic state represented by the vector (5) are not independent, they are, by definition, correlated.

Tensor products of vectors

The condition of independence just described can be expressed succinctly through the notion of a *tensor product*. Although tensor products are a very general notion, and can be defined quite abstractly and applied to a variety of mathematical structures, we can adopt a simple and concrete definition in the case at hand.

Given two vectors

$$|\phi\rangle = \sum_{a\in\Sigma} \alpha_a |a\rangle \quad \text{and} \quad |\psi\rangle = \sum_{b\in\Gamma} \beta_b |b\rangle,$$

the tensor product $|\phi\rangle \otimes |\psi\rangle$ is the vector defined as

$$|\phi\rangle \otimes |\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} \alpha_a\beta_b |ab\rangle.$$

The entries of this new vector correspond to the elements of the Cartesian product $\Sigma \times \Gamma$, which are written as strings in the previous equation. Equivalently, the vector $|\pi\rangle = |\phi\rangle \otimes |\psi\rangle$ is defined by the equation

$$\langle ab|\pi\rangle = \langle a|\phi\rangle\langle b|\psi\rangle$$

being true for every $a \in \Sigma$ and $b \in \Gamma$.

We can now recast the condition for independence: for a joint system (X, Y) in a probabilistic state represented by a probability vector $|\pi\rangle$, the systems X and Y are independent if $|\pi\rangle$ is obtained by taking a tensor product

$$|\pi\rangle = |\phi\rangle \otimes |\psi\rangle$$

of probability vectors $|\phi\rangle$ and $|\psi\rangle$ on each of the subsystems X and Y. In this situation, $|\pi\rangle$ is said to be a *product state* or *product vector*.

We often omit the symbol $\otimes$ when taking the tensor product of kets, such as writing $|\phi\rangle|\psi\rangle$ rather than $|\phi\rangle \otimes |\psi\rangle$. This convention captures the idea that the tensor product is, in this context, the most natural or default way to take the product of two vectors. Although it is less common, the notation $|\phi \otimes \psi\rangle$ is also sometimes used.

When we use the alphabetical convention for ordering elements of Cartesian products, we obtain the following specification for the tensor product of two column vectors.

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} \otimes \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} \alpha_1\beta_1 \\ \vdots \\ \alpha_1\beta_k \\ \alpha_2\beta_1 \\ \vdots \\ \alpha_2\beta_k \\ \vdots \\ \alpha_m\beta_1 \\ \vdots \\ \alpha_m\beta_k \end{pmatrix}$$

As an important aside, notice the following expression for tensor products of standard basis vectors:

$$|a\rangle \otimes |b\rangle = |ab\rangle.$$

We could alternatively write (a, b) as an ordered pair, rather than a string, in which case we obtain $|a\rangle \otimes |b\rangle = |(a, b)\rangle$. It is, however, more common to omit the parentheses in this situation, instead writing $|a\rangle \otimes |b\rangle = |a, b\rangle$. This is typical in mathematics more generally; parentheses that don't add clarity or remove ambiguity are often simply omitted.

The tensor product of two vectors has the important property that it is *bilinear*, which means that it is linear in each of the two arguments separately, assuming that the other argument is fixed. This property can be expressed through these equations:

1. Linearity in the first argument:

$$\big(|\phi_1\rangle + |\phi_2\rangle\big) \otimes |\psi\rangle = |\phi_1\rangle \otimes |\psi\rangle + |\phi_2\rangle \otimes |\psi\rangle$$
$$\big(\alpha|\phi\rangle\big) \otimes |\psi\rangle = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big)$$

2. Linearity in the second argument:

$$|\phi\rangle \otimes \big(|\psi_1\rangle + |\psi_2\rangle\big) = |\phi\rangle \otimes |\psi_1\rangle + |\phi\rangle \otimes |\psi_2\rangle$$
$$|\phi\rangle \otimes \big(\alpha|\psi\rangle\big) = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big)$$

Considering the second equation in each of these pairs of equations, we see that scalars "float freely" within tensor products:

$$\big(\alpha|\phi\rangle\big) \otimes |\psi\rangle = |\phi\rangle \otimes \big(\alpha|\psi\rangle\big) = \alpha\big(|\phi\rangle \otimes |\psi\rangle\big).$$

There is therefore no ambiguity in simply writing $\alpha|\phi\rangle \otimes |\psi\rangle$, or alternatively $\alpha|\phi\rangle|\psi\rangle$ or $\alpha|\phi \otimes \psi\rangle$, to refer to this vector.

Independence and tensor products for three or more systems

The notions of independence and tensor products generalize straightforwardly to three or more systems. If $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are systems having classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$, respectively, then a probabilistic state of the combined system $(\mathsf{X}_{n-1}, \ldots, \mathsf{X}_0)$ is a *product state* if the associated probability vector takes the form

$$|\psi\rangle = |\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle$$

for probability vectors $|\phi_0\rangle, \ldots, |\phi_{n-1}\rangle$ describing probabilistic states of $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$. Here, the definition of the tensor product generalizes in a natural way: the vector

$$|\psi\rangle = |\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle$$

is defined by the equation

$$\langle a_{n-1} \cdots a_0 | \psi \rangle = \langle a_{n-1} | \phi_{n-1} \rangle \cdots \langle a_0 | \phi_0 \rangle$$

being true for every $a_0 \in \Sigma_0, \ldots a_{n-1} \in \Sigma_{n-1}$.

A different, but equivalent, way to define the tensor product of three or more vectors is recursively in terms of tensor products of two vectors:

$$|\phi_{n-1}\rangle \otimes \cdots \otimes |\phi_0\rangle = |\phi_{n-1}\rangle \otimes \big(|\phi_{n-2}\rangle \otimes \cdots \otimes |\phi_0\rangle\big).$$

Similar to the tensor product of just two vectors, the tensor product of three or more vectors is linear in each of the arguments individually, assuming that all other arguments are fixed. In this case it is said that the tensor product of three or more vectors is *multilinear*.

Like in the case of two systems, we could say that the systems $\mathsf{X}_0, \ldots, \mathsf{X}_{n-1}$ are *independent* when they are in a product state, but the term *mutually independent* is more precise. There happen to be other notions of independence for three or more systems, such as *pairwise independence*, that are both interesting and important — but not in the context of this course.

Generalizing the observation earlier concerning tensor products of standard basis vectors, for any positive integer n and any classical states $a_0, \ldots, a_{n-1}$, we have

$$|a_{n-1}\rangle \otimes \cdots \otimes |a_0\rangle = |a_{n-1} \cdots a_0\rangle.$$

Measurements of probabilistic states

Now let us move on to measurements of probabilistic states of multiple systems. By choosing to view multiple systems together as single systems, we immediately obtain a specification of how measurements must work for multiple systems — provided that *all* of the systems are measured.

For example, if the probabilistic state of two bits (X, Y) is described by the probability vector

$$\frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle,$$

then the outcome 00 — meaning 0 for the measurement of X and 0 for the measurement of Y — is obtained with probability $1/2$ and the outcome 11 is also obtained with probability $1/2$. In each case we update the probability vector description of our knowledge accordingly, so that the probabilistic state becomes $|00\rangle$ or $|11\rangle$, respectively.

We could, however, choose to measure not *every* system, but instead just some of the systems. This will result in a measurement outcome for each system that gets measured, and will also (in general) affect our knowledge of the remaining systems that we didn't measure.

To explain how this works, we'll focus on the case of two systems, one of which is measured. The more general situation — in which some proper subset of three or more systems is measured — effectively reduces to the case of two systems when we view the systems that are measured collectively as if they form one system and the systems that are not measured as if they form a second system.

To be precise, let's suppose that X and Y are systems whose classical state sets are Σ and Γ, respectively, and that the two systems together are in some probabilistic state. We'll consider what happens when we measure just X and do nothing to Y. The situation where just Y is measured and nothing happens to X is handled symmetrically.

First, we know that the probability to observe a particular classical state $a \in \Sigma$ when just X is measured must be consistent with the probabilities we would obtain under the assumption that Y was also measured. That is, we must have

$$\mathrm{Pr}(\mathsf{X} = a) = \sum_{b \in \Gamma} \mathrm{Pr}\big((\mathsf{X}, \mathsf{Y}) = (a, b)\big).$$

This is the formula for the so-called reduced (or marginal) probabilistic state of X alone.

This formula makes perfect sense at an intuitive level, in the sense that something very strange would have to happen for it to be wrong. If it were wrong, that would mean that measuring Y could somehow influence the probabilities associated with different outcomes of the measurement of X, irrespective of the actual outcome of the measurement of Y. If Y happened to be in a distant location, such as somewhere in another galaxy for instance, this would allow for faster-than-light signaling — which we reject based on our understanding of physics. Another way to understand this comes from the interpretation of probability as reflecting a degree of belief. The mere fact that someone else might decide to look at Y cannot change the classical state of X, so without any information about what they did or didn't see, one's beliefs about the state of X should not change as a result.

Now, given the assumption that only X is measured and Y is not, there may still exist uncertainty about the classical state of Y. For this reason, rather than updating our description of the probabilistic state of (X, Y) to $|ab\rangle$ for some selection of $a \in \Sigma$ and $b \in \Gamma$, we must update our description so that this uncertainty about Y is properly reflected.

The following *conditional probability* formula reflects this uncertainty.

$$\mathrm{Pr}(\mathsf{Y} = b \,|\, \mathsf{X} = a) = \frac{\mathrm{Pr}\big((\mathsf{X}, \mathsf{Y}) = (a, b)\big)}{\mathrm{Pr}(\mathsf{X} = a)}$$

Here, the expression $\mathrm{Pr}(\mathsf{Y} = b \,|\, \mathsf{X} = a)$ denotes the probability that $\mathsf{Y} = b$ *conditioned* on (or *given* that) $\mathsf{X} = a$. Technically speaking, this expression only makes sense if $\mathrm{Pr}(\mathsf{X} = a)$ is nonzero, for if $\mathrm{Pr}(\mathsf{X} = a) = 0$, then we're dividing by zero and we obtain indeterminate form $\frac{0}{0}$. This is not a problem, though, because if the probability associated with a is zero, then we'll never obtain a as an outcome of a measurement of X, so we don't need to be concerned with this possibility.

To express these formulas in terms of probability vectors, consider a probability vector $|\psi\rangle$ describing a joint probabilistic state of (X, Y).

$$|\psi\rangle = \sum_{(a,b) \in \Sigma \times \Gamma} p_{ab} |ab\rangle$$

Measuring X alone yields each possible outcome $a \in \Sigma$ with probability

$$\mathrm{Pr}(\mathsf{X} = a) = \sum_{c \in \Gamma} p_{ac}.$$

The vector representing the probabilistic state of X alone is therefore given by

$$\sum_{a\in\Sigma}\left(\sum_{c\in\Gamma} p_{ac}\right)|a\rangle.$$

Having obtained a particular outcome $a \in \Sigma$ of the measurement of X, the probabilistic state of Y is updated according to the formula for conditional probabilities, so that it is represented by this probability vector:

$$|\pi_a\rangle = \frac{\sum_{b\in\Gamma} p_{ab}|b\rangle}{\sum_{c\in\Gamma} p_{ac}}.$$

In the event that the measurement of X resulted in the classical state a, we therefore update our description of the probabilistic state of the joint system (X,Y) to $|a\rangle \otimes |\pi_a\rangle$.

One way to think about this definition of $|\pi_a\rangle$ is to see it as a *normalization* of the vector $\sum_{b\in\Gamma} p_{ab}|b\rangle$, where we divide by the sum of the entries in this vector to obtain a probability vector. This normalization effectively accounts for a conditioning on the event that the measurement of X has resulted in the outcome a.

For a specific example, suppose that classical state set of X is $\Sigma = \{0,1\}$, the classical state set of Y is $\Gamma = \{1,2,3\}$, and the probabilistic state of (X,Y) is

$$|\psi\rangle = \frac{1}{2}|0,1\rangle + \frac{1}{12}|0,3\rangle + \frac{1}{12}|1,1\rangle + \frac{1}{6}|1,2\rangle + \frac{1}{6}|1,3\rangle.$$

Our goal will be to determine the probabilities of the two possible outcomes (0 and 1), and to calculate what the resulting probabilistic state of Y is for the two outcomes, assuming the system X is measured.

Using the bilinearity of the tensor product, and specifically the fact that it is linear in the *second* argument, we may rewrite the vector $|\psi\rangle$ as follows:

$$|\psi\rangle = |0\rangle \otimes \left(\frac{1}{2}|1\rangle + \frac{1}{12}|3\rangle\right) + |1\rangle \otimes \left(\frac{1}{12}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle\right).$$

In words, what we've done is to isolate the distinct standard basis vectors for the first system (that is, the one being measured), tensoring each with the linear combination of standard basis vectors for the second system we get by picking out the entries of the original vector that are consistent with the corresponding classical state of the first system. A

moment's thought reveals that this is always possible, regardless of what vector we started with.

Having expressed our probability vector in this way, the effects of measuring the first system become easy to analyze. The probabilities of the two outcomes can be obtained by summing the probabilities in parentheses.

$$\mathrm{Pr}(\mathsf{X} = 0) = \frac{1}{2} + \frac{1}{12} = \frac{7}{12}$$
$$\mathrm{Pr}(\mathsf{X} = 1) = \frac{1}{12} + \frac{1}{6} + \frac{1}{6} = \frac{5}{12}$$

These probabilities sum to one, as expected — but this is a useful check on our calculations.

And now, the probabilistic state of Y conditioned on each possible outcome can be inferred by normalizing the vectors in parentheses. That is, we divide these vectors by the associated probabilities we just calculated, so that they become probability vectors.

Thus, conditioned on X being 0, the probabilistic state of Y becomes

$$\frac{\frac{1}{2}|1\rangle + \frac{1}{12}|3\rangle}{\frac{7}{12}} = \frac{6}{7}|1\rangle + \frac{1}{7}|3\rangle,$$

and conditioned on the measurement of X being 1, the probabilistic state of Y becomes

$$\frac{\frac{1}{12}|1\rangle + \frac{1}{6}|2\rangle + \frac{1}{6}|3\rangle}{\frac{5}{12}} = \frac{1}{5}|1\rangle + \frac{2}{5}|2\rangle + \frac{2}{5}|3\rangle.$$

Operations on probabilistic states

To conclude this discussion of classical information for multiple systems, we'll consider *operations* on multiple systems in probabilistic states. Following the same idea as before, we can view multiple systems collectively as single, compound systems, and then look to the previous lesson to see how this works.

Returning to the typical set-up where we have two systems X and Y, let us consider classical operations on the compound system (X, Y). Based on the previous lesson and the discussion above, we conclude that any

such operation is represented by a stochastic matrix whose rows and columns are indexed by the Cartesian product $\Sigma \times \Gamma$.

For example, suppose that X and Y are bits, and consider an operation with the following description.

> If $\mathsf{X} = 1$, then perform a NOT operation on Y.
> Otherwise do nothing.

This is a deterministic operation known as a *controlled-NOT* operation, where X is the *control* bit that determines whether or not a NOT operation should be applied to the *target* bit Y. Here is the matrix representation of this operation:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Its action on standard basis states is as follows.

$$\begin{aligned} |00\rangle &\mapsto |00\rangle \\ |01\rangle &\mapsto |01\rangle \\ |10\rangle &\mapsto |11\rangle \\ |11\rangle &\mapsto |10\rangle \end{aligned}$$

If we were to exchange the roles of X and Y, taking Y to be the control bit and X to be the target bit, then the matrix representation of the operation would become

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

and its action on standard basis states would be like this:

$$\begin{aligned} |00\rangle &\mapsto |00\rangle \\ |01\rangle &\mapsto |11\rangle \\ |10\rangle &\mapsto |10\rangle \\ |11\rangle &\mapsto |01\rangle \end{aligned}$$

Another example is the operation having this description:

> Perform one of the following two operations, each with probability $1/2$:
>
> 1. Set Y to be equal to X.
> 2. Set X to be equal to Y.

The matrix representation of this operation is as follows:

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

The action of this operation on standard basis vectors is as follows:

$$\begin{aligned} |00\rangle &\mapsto |00\rangle \\ |01\rangle &\mapsto \frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle \\ |10\rangle &\mapsto \frac{1}{2}|00\rangle + \frac{1}{2}|11\rangle \\ |11\rangle &\mapsto |11\rangle \end{aligned}$$

In these examples, we are simply viewing two systems together as a single system and proceeding as in the previous lesson.

The same thing can be done for any number of systems. For example, imagine that we have three bits, and we increment the three bits modulo 8 — meaning that we think about the three bits as encoding a number between 0 and 7 using binary notation, add 1, and then take the remainder after dividing by 8. One way to express this operation is like this:

$$\begin{aligned} &|001\rangle\langle 000| + |010\rangle\langle 001| + |011\rangle\langle 010| + |100\rangle\langle 011| \\ &\quad + |101\rangle\langle 100| + |110\rangle\langle 101| + |111\rangle\langle 110| + |000\rangle\langle 111|. \end{aligned}$$

Another way to express it is as

$$\sum_{k=0}^{7} |(k+1) \bmod 8\rangle\langle k|,$$

assuming we've agreed that numbers from 0 to 7 inside of kets refer to the three-bit binary encodings of those numbers. A third option is to express this operation as a matrix.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Independent operations

Now suppose that we have multiple systems and we *independently* perform different operations on the systems separately.

For example, taking our usual set-up of two systems X and Y having classical state sets Σ and Γ, respectively, let us suppose that we perform one operation on X and, completely independently, another operation on Y. As we know from the previous lesson, these operations are represented by stochastic matrices — and to be precise, let us say that the operation on X is represented by the matrix M and the operation on Y is represented by the matrix N. Thus, the rows and columns of M have indices that are placed in correspondence with the elements of Σ and, likewise, the rows and columns of N correspond to the elements of Γ.

A natural question to ask is this: if we view X and Y together as a single, compound system (X, Y), what is the matrix that represents the combined action of the two operations on this compound system? To answer this question we must first introduce tensor products of matrices, which are similar to tensor products of vectors and are defined analogously.

Tensor products of matrices

The tensor product $M \otimes N$ of the matrices

$$M = \sum_{a,b \in \Sigma} \alpha_{ab} |a\rangle\langle b|$$

and

$$N = \sum_{c,d \in \Gamma} \beta_{cd} |c\rangle\langle d|$$

is the matrix

$$M \otimes N = \sum_{a,b \in \Sigma} \sum_{c,d \in \Gamma} \alpha_{ab} \beta_{cd} |ac\rangle\langle bd|$$

Equivalently, the tensor product of M and N is defined by the equation

$$\langle ac|M \otimes N|bd\rangle = \langle a|M|b\rangle \langle c|N|d\rangle$$

being true for every selection of $a, b \in \Sigma$ and $c, d \in \Gamma$.

An alternative, but equivalent, way to describe $M \otimes N$ is that it is the unique matrix that satisfies the equation

$$(M \otimes N)\big(|\phi\rangle \otimes |\psi\rangle\big) = \big(M|\phi\rangle\big) \otimes \big(N|\psi\rangle\big)$$

for every possible choice of vectors $|\phi\rangle$ and $|\psi\rangle$, assuming that the indices of $|\phi\rangle$ correspond to the elements of Σ and the indices of $|\psi\rangle$ correspond to Γ.

Following the convention described previously for ordering the elements of Cartesian products, we can also write the tensor product of two matrices explicitly as follows:

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mm} \end{pmatrix} \otimes \begin{pmatrix} \beta_{11} & \cdots & \beta_{1k} \\ \vdots & \ddots & \vdots \\ \beta_{k1} & \cdots & \beta_{kk} \end{pmatrix}$$

$$= \begin{pmatrix} \alpha_{11}\beta_{11} & \cdots & \alpha_{11}\beta_{1k} & & \alpha_{1m}\beta_{11} & \cdots \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots \\ \alpha_{11}\beta_{k1} & \cdots & \alpha_{11}\beta_{kk} & & \alpha_{1m}\beta_{k1} & \cdots \\ & \vdots & & \ddots & & \vdots \\ \alpha_{m1}\beta_{11} & \cdots & \alpha_{m1}\beta_{1k} & & \alpha_{mm}\beta_{11} & \cdots \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots \\ \alpha_{m1}\beta_{k1} & \cdots & \alpha_{m1}\beta_{kk} & & \alpha_{mm}\beta_{k1} & \cdots \end{pmatrix}$$

Tensor products of three or more matrices are defined in an analogous way. If $M_0, \ldots, M_{n-1}$ are matrices whose indices correspond to classical state sets $\Sigma_0, \ldots, \Sigma_{n-1}$, then the tensor product $M_{n-1} \otimes \cdots \otimes M_0$ is defined by the condition that

$$\langle a_{n-1} \cdots a_0 | M_{n-1} \otimes \cdots \otimes M_0 | b_{n-1} \cdots b_0 \rangle = \langle a_{n-1}|M_{n-1}|b_{n-1}\rangle \cdots$$

for every choice of classical states $a_0, b_0 \in \Sigma_0, \ldots, a_{n-1}, b_{n-1} \in \Sigma_{n-1}$. Alternatively, tensor products of three or more matrices can be defined recursively, in terms of tensor products of two matrices, similar to what we observed for vectors.

The tensor product of matrices is sometimes said to be *multiplicative* because the equation

$$(M_{n-1} \otimes \cdots \otimes M_0)(N_{n-1} \otimes \cdots \otimes N_0) = (M_{n-1}N_{n-1}) \otimes \cdots \otimes (M$$

is always true, for any choice of matrices $M_0, \ldots, M_{n-1}$ and $N_0 \ldots, N_{n-1}$, provided that the products $M_0N_0, \ldots, M_{n-1}N_{n-1}$ make sense.

Independent operations (continued)

We can now answer the question asked previously: if M is a probabilistic operation on X, N is a probabilistic operation on Y, and the two operations are performed independently, then the resulting operation on the compound system (X, Y) is the tensor product $M \otimes N$.

So, for both probabilistic states and probabilistic operations, *tensor products represent independence.* If we have two systems X and Y that are independently in the probabilistic states $|\phi\rangle$ and $|\pi\rangle$, then the compound system (X, Y) is in the probabilistic state $|\phi\rangle \otimes |\pi\rangle$; and if we apply probabilistic operations M and N to the two systems independently, then the resulting action on the compound system (X, Y) is described by the operation $M \otimes N$.

Let's take a look at an example, which recalls a probabilistic operation on a single bit from the previous lesson: if the classical state of the bit is 0, it is left alone; and if the classical state of the bit is 1, it is flipped to 0 with probability $1/2$. We observed that this operation is represented by the matrix

$$\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}.$$

If this operation is performed on a bit X, and a NOT operation is (independently) performed on a second bit Y, then the joint operation on the compound system (X, Y) has the matrix representation

$$\begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

By inspection, we see that this is a stochastic matrix. This will always be the case: the tensor product of two or more stochastic matrices is always stochastic.

A common situation that we encounter is one in which one operation is performed on one system and *nothing* is done to another. In such a case, exactly the same prescription is followed, bearing in mind that *doing nothing* is represented by the identity matrix. For example, resetting the bit X to the 0 state and doing nothing to Y yields the probabilistic (and in fact deterministic) operation on (X, Y) represented by the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Previous page Next page