# A First Look at Quantum Probability, Part 1

8-10 minutes

In this article and the next, I'd like to share some ideas from the world of quantum probability.* The word "quantum" is pretty loaded, but don't let that scare you. We'll take a *first*—not second or third—look at the subject, and the only prerequisites will be linear algebra and basic probability. In fact, I like to think of quantum probability as another name for "linear algebra + probability," so this mini-series will explore the mathematics, rather than the physics, of the subject.**

In today's post, we'll motivate the discussion by saying a few words about (classical) probability. In particular, let's spend a few moments thinking about the following:



What do I mean? We'll start with some basic definitions. Then I'll share an example that illustrates this idea.

A **probability distribution** (or simply, **distribution**) on a finite set $X$ is a function $p: X \to [0,1]$ satisfying $\sum_x p(x) = 1$. I'll use the term **joint probability distribution** to refer to a distribution on a Cartesian product of finite sets, i.e. a function $p: X \times Y \to [0,1]$ satisfying $\sum_{(x,y)} p(x,y) = 1$. Every joint distribution defines a **marginal probability distribution** on one of the sets by summing probabilities over the other set. For instance, the marginal distribution $p_X: X \to [0,1]$ on $X$ is defined by $p_X(x) = \sum_y p(x,y)$, in which the variable $y$ is summed, or "integrated," out. It's this very process of *summing* or *integrating out* that causes information to be lost. In other words, marginalizing loses information. It doesn't remember what was summed away!

I'll illustrate this with a simple example. To do so, I need to give you some finite sets $X$ and $Y$ and a probability distribution on them.

Consider the set $S$ of all bitstrings of length three. Every bitstring begins with either a 0 or a 1, so we can think of $S$ as the Cartesian product of the set $X = \{0, 1\}$ with the set of bitstrings of length two, $Y = \{00, 11, 01, 10\}$. It will be convenient to refer to elements in $X$ as **prefixes** and to elements in $Y$ as **suffixes**.
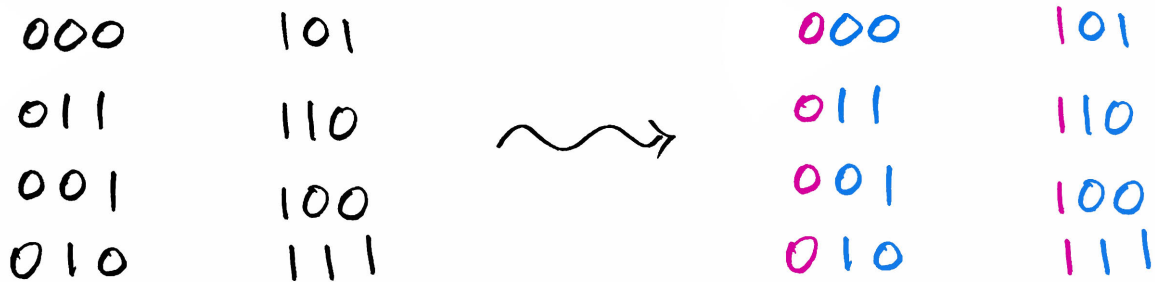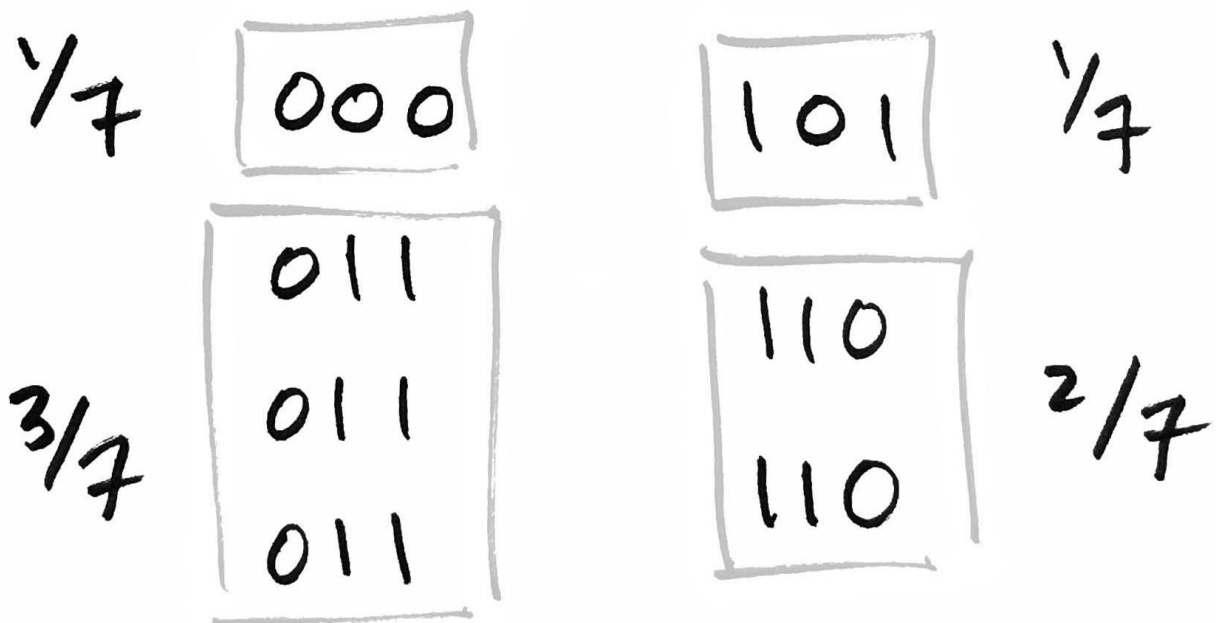
**prefixes**

$$X = \{0, 1\}$$

**suffixes**

$$y = \{00, 11, 01, 10\}$$

$$X \times Y \cong \text{all bitstrings of length 3:}$$

| | |
|---|---|
| 000 | 101 |
| 011 | 110 |
| 001 | 100 |
| 010 | 111 |

$\rightsquigarrow$

| | |
|---|---|
| 000 | 101 |
| 011 | 110 |
| 001 | 100 |
| 010 | 111 |

Now we need a probability distribution on $S \cong X \times Y$ to work with. Let's suppose the probability of 011 is $\frac{3}{7}$, the probability of 110 is $\frac{2}{7}$, the probability of both 000 and 101 is $\frac{1}{7}$, and the probability of the other bitstrings is zero. So you can imagine a seven-sided die with the faces labeled as follows:

$\frac{1}{7}$  | 000 |   | 101 | $\frac{1}{7}$

| 011 |

$\frac{3}{7}$  | 011 |

| 011 |

| 110 |

| 110 | $\frac{2}{7}$

One way to visualize this joint distribution is a weighted bipartite graph. We've explored this idea before. The set of prefixes and suffixes define the two sets of vertices (ergo, *bipartite*). An edge connects a prefix and a suffix if their concatenation is one of the samples drawn. That edge is labeled with the corresponding probability (ergo, *weighted*). Yet another way to view the joint distribution is as a $2 \times 4$ table. Rows are labeled by elements of $X$, columns are labeled by elements of $Y$, and the $xy$'th entry is the probability $p(x, y)$.

this joint distribution

can be viewed as a graph:
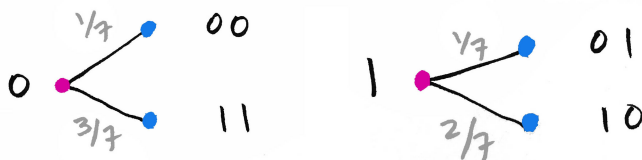
or as a table:

|   | 00 | 11 | 01 | 10 |
|---|-----|-----|-----|-----|
| 0 | $\frac{1}{7}$ | $\frac{3}{7}$ | 0 | 0 |
| 1 | 0 | 0 | $\frac{1}{7}$ | $\frac{2}{7}$ |

Marginal probabilities are easy to compute from a table: just sum along a row or column! For example, the probability of the suffix 00 is $p(00) = \frac{1}{7} + 0$.

marginal probabilities are sums along rows & columns:

|   | 00 | 11 | 01 | 10 |   |
|---|-----|-----|-----|-----|---|
| 0 | $\frac{1}{7}$ | $\frac{3}{7}$ | 0 | 0 | $\to p(0) = \frac{4}{7}$ |
| 1 | 0 | 0 | $\frac{1}{7}$ | $\frac{2}{7}$ | $\to p(1) = \frac{3}{7}$ |

$\downarrow$        $\downarrow$

$p(00) = \frac{1}{7}$     $p(10) = \frac{2}{7}$

Now here's the point I want to emphasize:

Marginal probability is forgetful.

- The marginal probability of the prefix 0 is $\frac{4}{7}$, but that number doesn't tell us that of the possible suffixes following 0, one is 00, three are 11, and none are 01 or 10.
- The marginal probability of the prefix 1 is $\frac{3}{7}$, but that number doesn't tell us that of the possible suffixes following 1, one is 01, two are 10, and none are 00 or 11.

In other words, summing over the suffixes in $Y$ has caused us to lose all information from $Y$ in a totally irreparable way. It cannot be recovered. There's no going back. This is what I meant by "marginal probability doesn't have memory." It's just a feature of probability.

*And this is where the fun begins.*

I'm now going to introduce a *different* way to compute the marginals from a joint distribution. It can be thought of as "marginal probability with memory." When computing marginal probabilities in this new way, you'll have ready access to the information lost in the old way!

The ideas are simple, though they might feel unmotivated at first. Bear with me. I want to show you something nice. Afterwards I'll explain what's going on.

## Here's a better way...

Again, we'll start with the joint distribution. As we know, it can be viewed as a $2 \times 4$ table. Here it is again:

$$M = \begin{bmatrix} \sqrt{1/7} & \sqrt{3/7} & 0 & 0 \\ 0 & 0 & \sqrt{1/7} & \sqrt{2/7} \end{bmatrix}$$

You see I've made some small cosmetic changes. The $2 \times 4$ table is now a $2 \times 4$ matrix, and I've judiciously added some square roots. You may ignore them if you like. I've included them for math reasons that we needn't worry about now. Lastly, I've given the matrix a name, $M$.

Let's now multiply $M$ by its transpose:

$$M^{\mathsf{T}}M = \begin{bmatrix} 1/7 & \sqrt{3}/7 & 0 & 0 \\ \sqrt{3}/7 & 3/7 & 0 & 0 \\ 0 & 0 & 1/7 & \sqrt{2}/7 \\ 0 & 0 & \sqrt{2}/7 & 2/7 \end{bmatrix}$$

This matrix is *very* interesting. For one, it's a $4 \times 4$ matrix. The set $Y$ has four elements in it and we can identify them with the rows/columns of this matrix. This sets the stage for another observation: **the diagonal of $M^{\mathsf{T}}M$ contains the marginal probability distribution on $Y$.**

the diagonals are the marginal probabilities on $Y$

$p(00) = 1/7$

$p(11) = 3/7$

$p(01) = 1/7$

$p(10) = 2/7$

$$\begin{array}{ccccc} & 00 & 11 & 01 & 10 \\ \begin{matrix} 00 \\ 11 \\ 01 \\ 10 \end{matrix} & \begin{bmatrix} 1/7 & \sqrt{3}/7 & 0 & 0 \\ \sqrt{3}/7 & 3/7 & 0 & 0 \\ 0 & 0 & 1/7 & \sqrt{2}/7 \\ 0 & 0 & \sqrt{2}/7 & 2/7 \end{bmatrix} \end{array}$$

So we've just computed marginal probability by multiplying a matrix by its transpose.

Voila.

*But wait, there's more.*

The off-diagonals of $M^\top M$ are also interesting. *Some are non-zero*. Let's not focus on their values for now (although they convey rich information). Let's just appreciate their existence: the fact that $M^\top M$ has non-zero off-diagonals means that it has interesting eigenvectors. It's a rank 2 matrix, so it has two eigenvectors. Here they are:

the eigenvectors of $M^+M$ :

$$
\begin{bmatrix} \sqrt{1/4} \\ \sqrt{3/4} \\ 0 \\ 0 \end{bmatrix}
$$

$\leftarrow \quad \frac{1}{4} \quad = \quad p(\text{suffix is } 00 \mid \text{prefix is } 0)$

$\leftarrow \quad \frac{3}{4} \quad = \quad p(\text{suffix is } 11 \mid \text{prefix is } 0)$

$\leftarrow \quad 0 \quad = \quad p(\text{suffix is } 01 \mid \text{prefix is } 0)$

$\leftarrow \quad 0 \quad = \quad p(\text{suffix is } 10 \mid \text{prefix is } 0)$

&

$$
\begin{bmatrix} 0 \\ 0 \\ \sqrt{1/3} \\ \sqrt{2/3} \end{bmatrix}
$$

$\leftarrow \quad 0 \quad = \quad p(\text{suffix is } 00 \mid \text{prefix is } 1)$

$\leftarrow \quad 0 \quad = \quad p(\text{suffix is } 11 \mid \text{prefix is } 1)$

$\leftarrow \quad \frac{1}{3} \quad = \quad p(\text{suffix is } 01 \mid \text{prefix is } 1)$

$\leftarrow \quad \frac{2}{3} \quad = \quad p(\text{suffix is } 10 \mid \text{prefix is } 1)$

Interesting indeed! **The square of the entries of these eigenvectors define conditional probability distributions on the set of suffixes, $Y$.** For instance, the first eigenvector defines a probability distribution on $Y$, conditioned on 0 being the prefix of a bitstring. Concretely: given that 0 is the prefix of a bitstring $s = (x, y) \in X \times Y$, the suffix $y$ will be 00

with probability $\left(\sqrt{\frac{1}{4}}\right)^2 = \frac{1}{4}$, it will be 11 with probability $\left(\sqrt{\frac{3}{4}}\right)^2 = \frac{3}{4}$, and it will be 01 and 10 each with probability $0^2 = 0$. This is the information contained in the entries of the first eigenvector. The second eigenvector likewise defines a probability distribution on $Y$, conditioned on the prefix $x = 1$.

> The information contained in the eigenvectors is precisely the information that's destroyed after computing marginal probability in the usual way.

But we're not done! We multiplied $M$ by its transpose on the left. If you switch the order, you'll see that the $2 \times 2$ matrix $MM^\top$ has the marginal distribution on the set of prefixes $X$ along its diagonal, and that its eigenvectors define conditional probabilities on $X$ after squaring the entries.

So $M^\top M$ has the marginals of $Y$ along its diagonal. And the existence of non-zero off-diagonals imply that its eigenvectors contain conditional probabilistic information; $MM^\top$ contains the same for $X$. Together, these two matrices recover all the information from the original joint distribution—something that can't be done when marginalizing in the usual way.

*Excellent.*

So, what's going on?

> The matrices $M^\top M$ and $MM^\top$ are the quantum versions of marginal probability distributions.

*What does that mean?*

I'll explain next time.

But here's a sneak peak:

The quantum version of a probability distribution is something called a *density operator*. The quantum version of marginalizing corresponds to "reducing" that operator to a subsystem. This reduction is a construction in linear algebra called the partial trace. In Part 2 of this miniseries, I'll begin by explaining the partial trace. Then I'll explain what I mean by "quantum version." Along the way, we'll unwind the basics of quantum probability theory.

Until then!