

# Chapter 1

## Quantum Measurement Theory

Before we begin, it is worth mentioning a few things about terminology. In mathematics texts it is usual to denote a random variable as a capital letter, say  $X$ , and the variable denoting one of the values it can take as the corresponding lower case letter,  $x$ . This provides technical precision, since the concept of a random variable, and the concept of one of the values it can take, are distinct. However, physicists tend to use the same symbol to denote both things, because it causes no confusion in practice. This is the style I prefer, so I will use it here.

Secondly, we physicists often use the term “probability distribution” as synonymous with “probability density”, whereas mathematicians use the former term to mean the anti-derivative of a probability density. Defining “distribution” to be synonymous with “density” is convenient, because “probability density” can only be used to refer to continuous distributions and not to discrete ones, whereas “probability distribution” can be used for both. For this reason we will use the term *probability distribution* to mean probability density.

This text is for graduate physics students, and as such we assume that the reader is familiar with quantum mechanics, the basics of probability theory, and various mathematical concepts such as Fourier transforms and  $\delta$ -functions. Everything that one needs to know about probability theory and Fourier transforms can be found in Chapter 1 of reference [1] or Chapter 4 of reference [2] and Chapter 1 of reference [3]. We also recommend Jaynes’ excellent book on probability theory [4].

### 1.1 Classical Measurement Theory

Classical measurement theory, also known as Bayesian statistical inference, tells us how our knowledge about the value of some quantity,  $x$ , changes when we obtain a piece of data relating to  $x$ . To understand how this works, we need to know first how to describe the knowledge we have regarding  $x$ . We will assume that  $x$  is some variable that can take any real value, but more generally it could be a vector of discrete or continuous variables. Since we need to make a measurement to determine  $x$ , we must be uncertain about its value. Our knowledge about  $x$  is therefore captured by a

probability distribution,  $P(x)$ , for the values of  $x$ . This probability distribution tells us, based on the information currently available, the likelihood that  $x$  will have various values, and overall how certain, or uncertain, we are about  $x$ . This distribution is called our *state-of-knowledge* of  $x$ .

To determine how our state-of-knowledge changes when we obtain a piece of data  $y$ , we have to know how  $y$  is related to  $x$ . To be able to describe measurements, this relationship must be probabilistic: if  $y$  were deterministically related to  $x$  (meaning that  $y$  was a function of  $x$ ), then we could determine  $x$  precisely once we knew  $y$ , simply by inverting the function. This is not what happens when we make measurements; measurements are never perfectly precise. After we have made a measurement of  $x$ , giving a result  $y$ , we are always left with some uncertainty about  $x$ .

Consider measuring the length of an object with a ruler. In this case the result of the measurement is equal to the true length plus a random error. Thus, given the true length of the object (the value of  $x$ ) there is a probability distribution for the result  $y$ . The probability distribution for  $y$  is peaked at  $y = x$ , and thus the probability distribution for  $y$  is a function of  $x$ . Because this probability distribution for  $y$  depends on (is conditional upon) the value of  $x$ , it is called a *conditional* probability distribution for  $y$ , and is written  $P(y|x)$  (this is read as ‘‘P of  $y$  given  $x$ ’’). The conditional probability for the measurement result (the data)  $y$ , given the value of the quantity to be measured,  $x$ , completely defines the measurement. This conditional probability is determined by the physical procedure used to make the measurement, and is referred to as the *likelihood function* for the measurement.

To determine how our state-of-knowledge regarding  $x$ ,  $P(x)$ , changes when we obtain the value of  $y$ , we use the relationships between the joint probability for two random variables  $x$  and  $y$ ,  $P(x, y)$ , and the conditional probabilities  $P(y|x)$  and  $P(x|y)$ . These relationships are

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x). \quad (1.1)$$

Here  $P(y)$  is the probability distribution for  $y$  irrespective of the value of  $x$  (also called the *marginal* distribution for  $y$ ), and is given by

$$P(y) = \int_{-\infty}^{\infty} P(x, y)dx. \quad (1.2)$$

Now  $P(x)$  is our state-of-knowledge of  $x$  prior to making the measurement, and it is therefore the probability density for  $x$  irrespective of the value of  $y$ . It is therefore also the marginal probability for  $x$ , and thus given by

$$P(x) = \int_{-\infty}^{\infty} P(x, y)dy. \quad (1.3)$$

While the relationships in Eq.(1.1) are fairly intuitive, they are explained further in, for example, references [1] and [2].

Rearranging Eq.(1.1) we obtain the famous relationship known as *Bayes' Theorem*, being

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}. \quad (1.4)$$

Upon examining Eq.(1.4) we will see that it tells us exactly how to change our state-of-knowledge when we obtain the measurement result  $y$ . First note that since  $P(x|y)$  must be normalized (that is, its integral over  $x$  must be unity), the value of  $P(y)$  on the bottom line is completely determined by this normalization. We can therefore write Bayes' theorem as

$$P(x|y) = \frac{P(y|x)P(x)}{\mathcal{N}}, \quad \text{where } \mathcal{N} = \int_{-\infty}^{\infty} P(y|x)P(x)dx = P(y). \quad (1.5)$$

We see that on the right hand side (RHS) we have our state-of-knowledge of  $x$  before we obtain the data  $y$ , and on the LHS the probability for  $x$  given that value of  $y$ . The LHS is therefore our state-of-knowledge after obtaining the value of  $y$ . In Eq.(1.5)  $P(x)$  is called the *prior* probability, and  $P(x|y)$  the *posterior* probability. So Bayes' theorem tells us that to obtain our new state-of-knowledge once we have made our measurement: we simply multiply our current state-of-knowledge by the likelihood function  $P(y|x)$ , and normalize the result. Note that the prior is simply the marginal (overall) distribution of  $x$ . The relationship given by Eq.(1.5), Bayes' Theorem, is the fundamental theorem of classical measurement theory.

### 1.1.1 Understanding Bayes' theorem

While Bayes' theorem (Eq.(1.5)) is simple to derive, to obtain a direct understanding of it requires a bit more work. To this end consider a measurement of a discrete variable,  $x$ , in which  $x$  has only two values. We will call these values  $\mathbf{0}$  and  $\mathbf{1}$ . Our measurement will also have only two outcomes, which we will denote by  $y = 0$  and  $y = 1$ . In our discussion of Bayes' theorem above, we assumed that  $x$  was continuous, so let us take a minute to re-orient our thinking to discrete variables. In the present case our state-of-knowledge,  $P(x)$ , has only two values, being  $P(\mathbf{0})$  and  $P(\mathbf{1})$  (and, of course,  $P(\mathbf{0}) + P(\mathbf{1}) = 1$ ). The conditional probability,  $P(y|x)$ , also has only two values for each value of  $x$ . If we make the measurement and obtain the result  $y = 1$  (for example), then Bayes' theorem tells us that our posterior is given by

$$P(x|1) = \frac{P(1|x)P(x)}{\sum_{x'} P(1|x')P(x')} = \frac{P(1|x)P(x)}{P(1|\mathbf{0})P(\mathbf{0}) + P(1|\mathbf{1})P(\mathbf{1})}. \quad (1.6)$$

We now wish to obtain a better understanding of this expression.

To do so let us choose a specific likelihood function for the measurement. This likelihood function is given in table 1.1, and contains the parameter  $\alpha$ . If  $\alpha$  is close to unity, then the two values of  $x$  give very different distributions for the measurement result  $y$ , and in this case we would expect the measurement to tell us a lot about  $x$ . Conversely, if  $\alpha$  is close to  $1/2$ , then the reverse is true. For the sake of concreteness

$P(y x)$	$y = 0$	$y = 1$
$x = 0$	$\alpha$	$1 - \alpha$
$x = 1$	$1 - \alpha$	$\alpha$

Table 1.1: The likelihood function for a simple “two-outcome” measurement.

let us choose  $0.5 < \alpha < 1$ . With this choice, if  $x = \mathbf{0}$  then it is more likely that we will get the result  $y = 0$ , and if  $x = \mathbf{1}$  it is more likely that we will get  $y = 1$ .

Now assume that we initially know nothing about  $x$ , so that our prior state of knowledge is  $P(\mathbf{0}) = 0.5 = P(\mathbf{1})$ . What happens when we make the measurement and get the result  $y = 1$ ? Since our prior is uniform, Bayes’ theorem tells us that our posterior is simply the likelihood function (normalized if necessary), and in this case it is already normalized. Our posterior is thus  $P(\mathbf{0}|1) = 1 - \alpha$  and  $P(\mathbf{1}|1) = \alpha$ . So it is now more likely that  $x = \mathbf{1}$  than  $x = \mathbf{0}$ . This is indeed intuitively reasonable. The likelihood function tells us that if  $x$  were to be  $\mathbf{0}$  then it would be less likely that we would get the result 1, so it is reasonable that since we obtained the result 1, it is more likely that  $x = \mathbf{1}$  than that  $x = \mathbf{0}$ .

The above discussion shows that if the prior is uniform, Bayes’ theorem tells us that *the values of  $x$  that are more likely are those for which the result we have obtained is the more likely outcome.*

Now let us examine why we need to include the prior, in addition to the likelihood function, when calculating our posterior. This is clearest when the prior is strongly weighted towards one value of  $x$ . Consider a situation in which the prior is  $P(\mathbf{0}) = 0.999$  and  $P(\mathbf{1}) = 0.001$ . This means that in the absence of any further data,  $x$  will only be equal to  $\mathbf{1}$  one time in a thousand cases. Now let us consider a slightly different two-come measurement from the one above (this new choice will make the logic simple). The likelihood function for this new measurement is given in table 1.2, and in words is as follows: if  $x = \mathbf{0}$  then  $y$  is always equal to 0. If  $x = \mathbf{1}$ , then  $y = 1$  with probability 0.999 and  $y = 0$  only one time in a thousand. This means that, if the prior is flat, upon obtaining the result  $\alpha = 1$  the value of  $x$  would be equal to  $\mathbf{1}$  nine hundred and ninety nine times out of thousand. So what is the case when the prior is as above, highly weighted towards  $x = \mathbf{0}$ ? First we note that  $x$  is equal to  $\mathbf{1}$  only one time in a thousand. Now, if we get the result  $y = 1$  there are two possibilities. Either  $x = \mathbf{1}$ , which happens one time in one thousand, or  $x = \mathbf{0}$  and we go the result  $y = 1$  anyway, which also happens approximately one time in a thousand. (The precise figure is the frequency of  $x = \mathbf{0}$  multiplied by the frequency of the result  $y = 1$  given  $x = \mathbf{0}$ , which is  $0.999 \times 0.001 \approx 1/(1001)$ ). Thus the result  $y = 1$  happens approximately one time in 500, and half of these are due to  $x = \mathbf{0}$ , and half due to  $x = \mathbf{1}$ . Thus, when we obtain the result  $y = 1$ , there is only a 50% chance that  $x = \mathbf{1}$ . This, is of course, exactly what Bayes’ theorem tells us; by multiplying the likelihood function that weights  $x = \mathbf{1}$  very highly, by the prior which weights  $x = \mathbf{0}$  very highly, we obtain an approximately flat posterior.

$P(y x)$	$y = 0$	$y = 1$
$x = 0$	1	0
$x = 1$	0.001	0.999

Table 1.2: The likelihood function for our second example of a simple “two-outcome” measurement.

The example we have just considered applies directly to a real, and very important situation: testing for the HIV virus. Each test is pretty reliable, giving a false positive only about one time in a thousand. On that basis alone one might think that when a result comes back positive, there is little reason to perform a follow-up test to confirm it. But this is very wrong. Since very few patients have HIV, false positives come up just as frequently as real positives. Thus, whenever a positive test result comes back, it is essential to do a follow-up test to check that it is not a false positive. Bayesian inference, and thus measurement theory, is therefore crucial in real world problems.

To complete our discussion of Bayes’ theorem it is worth noting that our state-of-knowledge does not necessarily become more certain when we make a measurement. To take the example of the HIV test above, before we obtain the result of the test we are almost certain that the patient is HIV negative, since the vast majority of patients are. However, upon obtaining a positive test result, there is an approximately fifty-fifty chance that the patient is HIV positive. Thus, after obtaining the measurement result, we are less certain of the HIV status of the patient. Even in view of this, all measurements do have the property that, upon making the measurement, we become more certain *on average* (where the average is taken over all the possible measurement results). We will be able to make this statement precise once we have described how to quantify the concept of information in Chapter 2.

## 1.1.2 Multiple measurements and Gaussian distributions

### Multiple measurements

Having made a measurement of  $x$ , what happens when we make a second measurement? We might expect that we simply repeat the process of multiplying our current state-of-knowledge, now given by  $P(x|y)$ , by the likelihood function for the new measurement. This is correct so long as the results of the two measurements are independent, and is simple to show as follows. Let us say that we make  $N$  measurements, and the results (the data) obtained from these measurements are  $y_i$  for  $i = 1, \dots, N$ . Bayes’ theorem tells us that

$$P(x|y_1, \dots, y_N) = \frac{P(y_1, \dots, y_N|x)P(x)}{\mathcal{N}}, \quad (1.7)$$

with  $\mathcal{N} = \int_{-\infty}^{\infty} P(y_1, \dots, y_N | x) P(x) dx$ . The fact that all the measurement results are independent means that

$$P(y_1, \dots, y_N | x) = P(y_1 | x) P(y_2 | x) \cdots P(y_N | x), \quad (1.8)$$

and with this Bayes' theorem becomes

$$\begin{aligned} P(x | y_1, \dots, y_N) &= \frac{P(y_1, \dots, y_N | x) P(x)}{\mathcal{N}} \\ &= \frac{P(y_N | x) \cdots P(y_1 | x) P(x)}{\mathcal{N}} \\ &= \frac{P(y_N | x)}{\mathcal{N}_N} \cdots \frac{P(y_1 | x) P(x)}{\mathcal{N}_1}. \end{aligned} \quad (1.9)$$

So we see that each time we make another independent measurement we update our state-of-knowledge by multiplying it by the likelihood function and normalizing the result.

### Pooling independent knowledge

It turns out that classical measurement theory provides us with a simple way to pool the knowledge of two observers so long as their information has been obtained independently. If two observers, A and B, have the respective states-of-knowledge  $P_1(x)$  and  $P_2(x)$  about a quantity  $x$ , then we can write each of these as

$$P_A(x) = \frac{P(\mathbf{y}_A | x) P_{\text{prior}}(x)}{\mathcal{N}_A}, \quad (1.10)$$

$$P_B(x) = \frac{P(\mathbf{y}_B | x) P_{\text{prior}}(x)}{\mathcal{N}_B}, \quad (1.11)$$

where  $\mathbf{y}_A$  and  $\mathbf{y}_B$  are the vectors of data obtained by the respective observers. Since we intend the data the observers have obtained to represent *all* the information they each have about  $x$ ,  $P_{\text{prior}}(x)$  is the prior that describes no knowledge about  $x$ . If we choose the measure for integration over  $x$  so that the prior corresponding to knowing nothing is equal to unity ( $P_{\text{prior}}(x) = 1$ ), then an observer who has access to the data of both A and B has the state-of-knowledge

$$P(x) = \frac{P(\mathbf{y}_A | x) P(\mathbf{y}_B | x)}{\mathcal{N}'} = \frac{P_A(x) P_B(x)}{\mathcal{N}}. \quad (1.12)$$

So all we have to do to pool the knowledge of two (or more) observers is to multiply their states-of-knowledge together, and normalize the result.

### The ubiquity of Gaussian measurements

Now consider applying classical measurement theory to the simple example discussed above, that of measuring the length of an object with a ruler. To describe this

measurement we need to decide how the result that we get,  $y$ , depends upon the true length  $x$ . We can think of  $y$  as being equal to the true length plus a random “error”. The error in a measurement is often described well by a Gaussian distribution. The reason for this is that the error is usually the result of random contributions from many different sources. When we add many independent random variables together, the central limit theorem tells us that the resulting random variable has an approximately Gaussian distribution. If the error in our measurement of  $x$  is a Gaussian with mean zero and variance  $V$ , then the probability distribution for  $y$ , given  $x$ , is a Gaussian centered at  $x$ :

$$P(y|x) = \frac{1}{\sqrt{2\pi V}} e^{-(y-x)^2/(2V)}. \quad (1.13)$$

This is the likelihood function for the measurement. If we have absolutely no knowledge of  $x$  before the measurement (never the case in reality, of course), then we can set  $P(x) = 1$ . Our knowledge of  $x$  after making the measurement is then simply the likelihood function, normalized so that it is a valid probability distribution over  $x$  for each value of  $y$ . In this case the normalization is already correct, and so

$$P(x|y) = \frac{1}{\sqrt{2\pi V}} e^{-(x-y)^2/(2V)}. \quad (1.14)$$

This tells us that the value of  $x$  is a Gaussian centered at  $y$ , with variance  $V$ , and thus that the most likely value of  $x$  is  $y$ , and the expectation value of  $x$  is also equal to  $y$ . It is customary to quote the error on a measured quantity as twice the standard deviation. Thus we write the value of  $x$  as  $y \pm 2\sqrt{V}$ .

### The ubiquity of Gaussian states-of-knowledge

It is not only measurement errors that tend to be Gaussian, but also states-of-knowledge. The reason for this is that when one makes many measurements of a given quantity, the resulting state-of-knowledge is the result of multiplying together many likelihood functions. One can show that even if these likelihood functions are not themselves Gaussian, the result usually is. This is because there is a central limit theorem for multiplication that follows from the usual central limit theorem for summing random variables. To see how this works we first recall that if we have two independent random variables  $x$  and  $y$ , the probability distribution for their sum,  $z = x + y$ , is the *convolution* of the distributions for  $x$  and  $y$ :

$$P_z(z) = P_x(x) * P_y(y) \equiv \int_{-\infty}^{\infty} P_x(u)P_y(z-u)du. \quad (1.15)$$

The central limit theorem states that so long as the variances of all the convolved distributions are finite, then as one convolves more distributions together, the closer the result is to a Gaussian.

To obtain a similar result for multiplying a number of probability distributions together, we take the Fourier transforms of these probability distributions. Multiplying two functions together is the same as convolving their Fourier transforms. That is, the Fourier transform of the product of two functions is the convolution of their Fourier transforms. So the central limit theorem tells us that if we convolve the Fourier transforms of a set of distributions together, then the result is close to a Gaussian. Since the inverse Fourier transform of this approximate Gaussian is the product of all the probability distributions, and since the Fourier transform of a Gaussian is also Gaussian, the product of the distributions is approximately Gaussian.

So when we make multiple measurements to learn about some quantity, or when our knowledge about a quantity is derived from many different sources, then our state-of-knowledge will be approximately Gaussian. Just as Gaussian errors are common in nature, Gaussian states-of-knowledge are also common.

Adding noise and making measurements can be regarded as opposite processes. The more noise (random variables) we add to a quantity, the more uncertain it becomes. The more measurements we make on a quantity, the more certain (on average) it becomes. It is interesting to note that the mathematical operations that describe each of these processes, convolution for adding noise, and multiplication for making measurements, are dual to each other from the point of view of the Fourier transform.

### 1.1.3 Prior states-of-knowledge and invariance

Bayes' theorem makes it clear how to update our state-of-knowledge whenever we obtain new data. So long as we know the overall distribution for the unknown quantity (the prior), then Bayes' theorem is strictly correct and open to no interpretation. However, for many situations it is not clear what the prior is. Before we discuss the main tool for determining priors, it is important to note that, in many situations, the choice of prior is, in fact, unimportant. This is the case when we can obtain enough data that the likelihood function is much sharper (has much less uncertainty) than any reasonable prior for the problem. In this case the prior has very little influence on the final expectation value of the measured quantity, and is therefore unimportant. In this case we can ignore the prior completely and set the posterior equal to the (correctly normalized) likelihood function for the set of measurements.

In many problems in which we need to specify a prior, we need the prior to capture what it means for us to "know nothing" about the quantity to be measured. The reason for this is that we do not want the prior to bias our results one way or the other; the prior should not contain any hidden assumptions of prior prejudice. The question of what it means to know nothing is actually quite subtle. In the case of measuring the position of an object in space, it seems quite obvious that the prior should be constant and extend from  $-\infty$  to  $\infty$ . This prior is not normalizable, but this does not matter, since the likelihood function, and thus the posterior, will be for any reasonable measurement. However, the question of what it means to know



nothing about a quantity that can only be positive, such as the number of sand grains in a sand pile, is not at all obvious. As will become clear, a prior that is constant on the interval  $(1, \infty)$  is not the correct choice for this case.

The problem of what it means to know nothing about a positive quantity is solved by using the idea that the prior should be invariant under a transformation. The key is to realize that if one knows nothing about the value of a positive quantity,  $\lambda$ , then multiplying  $\lambda$  by any number should not change one's state-of-knowledge. That is, if one is completely ignorant of  $\lambda$ , then one must be ignorant of the overall *scale* of  $\lambda$ . There is only one state-of-knowledge that is invariant under a scaling transformation, and that is

$$P(\lambda) = \frac{1}{\lambda}. \quad (1.16)$$

This prior is also not normalizable, but this is fine so long as the likelihood function is normalizable.

The powerful technique of identifying transformations over which our prior state-of-knowledge must be invariant was introduced by Jaynes [5, 4], inspired by the work of Jeffreys [6, 7, 8]. Returning to the problem of the prior for the location of an object, we see that the flat prior is the one that is invariant under all translations, being the natural transformation in that case. Natural transformations for a given quantity are those transformations that map all admissible values of the quantity to admissible values. That is why the translation transformation is not appropriate for quantities that can only be positive. An excellent example of the use of invariance for determining priors is given by Jaynes' solution to "Bertrand's problem", and can be found in [4].

Using the notion of invariance it is simple to determine priors for quantities that have symmetries. The simplest case is a "six-sided" die. If the die is perfectly symmetric, then our prior should be invariant under all permutations of the six-faces of the die. The result is a prior that assigns equal probability to all the faces. A less trivial example is a prior over the possible states of an  $N$ -dimensional quantum system. These states live in an  $N$ -dimensional complex vector space. In this case the natural set of transformations are those given by all the unitary transformations on the vector space. In this case the zero-knowledge prior is the one that is invariant under all unitaries. Actually what one does for this problem is to choose the measure over which to integrate so that the probability distribution is unity, and this means selecting the measure that is unitarily covariant. This measure is called the *Harr measure*, and is given in Appendix C. The definition of a measure, using a minimum of technical jargon, can be found in reference [1].

### **Priors: a tricky example**

For the sake of interest we pause to consider an example that shows the kinds of problems that can arise when choosing priors. This is the apparent paradox of the "two-envelopes problem". In this problem we are given two envelopes, one of which contains twice as much money as the other. We are allowed to open only one of

the envelopes, and then we get to choose which envelope we will take home to spend. Before we open one of the envelopes, clearly there is no reason to favor either envelope. Assume that we now open one of the envelopes and find that it contains  $m$  dollars. We now know that the second envelope (the one we haven't opened) contains either  $m/2$  dollars or  $2m$  dollars. We conclude that if we choose to keep the second envelope instead of the first, then we stand to lose  $m/2$  dollars, or gain  $x$  dollars. We thus stand to gain more than we will lose by choosing the second envelope, and so it appears that we should do so. There is clearly something wrong with this reasoning, because if we had instead chosen to open the second envelope, then precisely the same reasoning would now tell us to take home the first envelope!

What is wrong with the above analysis is the hidden assumption we have made about what it means to initially “know nothing” about the amount of money in the envelopes. That is, we have arbitrarily chosen a prior, and this causes problems with our reasoning. Now consider the problem again, with the prior made explicit, and in a situation in which we can easily determine what the prior should be. As before we are given two envelopes, one of which contains twice as much money as the other, but this time we are told that the smaller amount of money has a uniform distribution between 0 and 100 dollars. This distribution is our prior. Now if we open one of the envelopes, and find that it contains 10 dollars we can work out precisely the relative probabilities that the second envelope contains either 5 dollars or 20 dollars, and make our choice appropriately. Since the prior is flat in the smaller amount, in this case we should definitely pick the second envelope. However, if we open the first envelope and find that it contains 110 dollars, then we know that we should keep it; in this case Bayesian inference tells us that since the smaller amount cannot be larger than 100 dollars, this envelope contains the larger amount. There is no paradox so long as we choose a meaningful prior. In the original statement of the problem, it is assumed implicitly that knowing nothing means that the prior is flat from zero out to infinity. This is obviously nonsense, because if this were really the case, when we opened the envelope we would find an unimaginably large sum of money in it — in fact, we would find an infinite amount of money in it. We know, therefore, that there must be some kind of upper limit in the prior.

The problem with the “two-envelopes” paradox is that we do not know how to quantify what it means to say that we know that one of the envelopes contains twice as much money as the other, but know nothing else about these amounts. We might assume that it should be the scale invariant prior for the smaller amount, but this cannot be the case because it does not resolve the paradox; apparently the scale-invariant prior does not take into account that there are two related quantities to consider, rather than only one. Nevertheless, what we do know is that *if* we knew nothing to begin with, then upon opening one of the envelopes we must still have no reason to prefer one envelope over the other. Thus the prior that captures what it means to know nothing in this case must be the one that tells us, upon opening the first envelope, that we stand to neither gain nor lose, on average, by switching

to the other envelope. This means that our state-of-knowledge regarding the second envelope, once we open the first, must be such that the average amount of money in the second envelope is equal to the amount in the first. It is not difficult to use this to determine the prior, and we leave it as an exercise.

## 1.2 Quantum Measurement Theory

### 1.2.1 The measurement postulate

The state of a quantum system,  $|\psi\rangle$ , is a vector in a complex vector space. If the set of vectors  $\{|n\rangle\}$ ,  $n = 0, \dots, N - 1$  (where  $N$  may be  $\infty$ ) is an orthonormal basis for this space, then we can always express  $|\psi\rangle$  as

$$|\psi\rangle = \sum_n c_n |n\rangle \quad (1.17)$$

for some complex coefficients  $c_n$ , where  $\sum_n |c_n|^2 = 1$ .

The basis of quantum measurement theory is the following postulate: We can choose any basis, and look to see which one of these basis states the system is in. When we do so, we *will* find the system to be in one of these basis states, even though it may have been in *any* state  $|\psi\rangle$  before the measurement. Which basis state we find is random. If the system is initially in the state  $|\psi\rangle$  then the probability that we will find state  $|n\rangle$  is given by  $|c_n|^2$ . A measurement like this, for which the result is one of a set of basis states, is called a *von Neumann* measurement. Before we use this basic measurement postulate to derive quantum measurement theory (which has, necessarily, a very similar structure to classical measurement theory), we need to know how to describe states-of-knowledge about quantum systems.

### 1.2.2 Quantum states-of-knowledge: density matrices

Quantum states already contain within them probabilities — once we express a quantum state in some basis, the coefficients for that basis determine the probabilities for finding the system in those basis states. However, these probabilities are not enough to describe all the possible states-of-knowledge that we might have about a quantum system. Even though a system may actually *be* in a given state  $|\psi\rangle$ , we may not know what this state is. In general then, our state-of-knowledge about a quantum system can be described by a probability density over all the possible states  $|\psi\rangle$ . We might refer to this probability density as describing our *classical* uncertainty about the system, and the coefficients  $c_n$  as describing the quantum uncertainty inherent in a given quantum state-vector.

While a complete state-of-knowledge of a quantum system is a probability density over all the possible states  $|\psi\rangle$ , for most purposes one can use a more compact representation of this state-of-knowledge. This compact representation is called the

*density matrix*. It was devised independently by von Neumann [9] and Landau [10] in 1927.

To obtain the density matrix formalism, we first recall that the expectation value of a physical observable, for a system in state  $|\psi\rangle$ , is given by

$$\langle X \rangle = \langle \psi | X | \psi \rangle = \text{Tr}[X |\psi\rangle\langle\psi|], \quad (1.18)$$

where  $X$  is the operator corresponding to that observable. So while the expectation value of an observable is quadratic in the vector  $|\psi\rangle$ , it is linear in the operator (matrix)  $|\psi\rangle\langle\psi|$ . Expectation values are also linear in classical probabilities. If our state-of-knowledge is a probability distribution over the  $M$  states  $\{|\phi_m\rangle\}$ , where the probability of the system being in the state labeled by  $m$  is  $p_m$ , then the expectation value of  $X$  is

$$\begin{aligned} \langle X \rangle &= \sum_m p_m \langle \phi_m | X | \phi_m \rangle = \sum_m p_m \text{Tr}[X |\phi_m\rangle\langle\phi_m|] \\ &= \text{Tr} \left[ X \left( \sum_m p_m |\phi_m\rangle\langle\phi_m| \right) \right]. \end{aligned} \quad (1.19)$$

So we see that the matrix

$$\rho \equiv \sum_m p_m |\phi_m\rangle\langle\phi_m| \quad (1.20)$$

is sufficient to calculate the expectation value of any operator. We will see below that it is also sufficient to calculate the results of any measurement performed on the system. Note that since it is only the results of measurements on quantum systems that determine events in the macroscopic world (if you are not convinced of this assertion, a more detailed explanation is given in section 4.2.3), questions about the future behavior of a quantum system are ultimately questions about the results of measurements. Thus  $\rho$  is sufficient to fully characterize the future behavior of a quantum system, and this is why it is a sufficient description of one's state-of-knowledge for many purposes.

In the absence of measurements the evolution of  $\rho$  is very simple. This evolution is simply given by evolving each of its component states. The evolution of a quantum state is given by applying to it a unitary operator:  $|\psi(t)\rangle = U|\psi(0)\rangle$ . So the evolution of  $\rho$  is

$$\begin{aligned} \rho(t) &= \sum_m p_m |\phi_m(t)\rangle\langle\phi_m(t)| = \sum_m p_m U |\phi_m(0)\rangle\langle\phi_m(0)| U^\dagger \\ &= U \rho(0) U^\dagger. \end{aligned} \quad (1.21)$$

The “density matrix”  $\rho$  is therefore a simple and compact way to describe our state-of-knowledge of a quantum system. The term *density matrix* comes from its role as a quantum equivalent of a probability density for classical systems, and also because

its diagonal elements constitute a probability distribution. Specifically, if we express our states in the basis  $\{|j\rangle\}$ , so that

$$|\phi_m\rangle = \sum_j c_{jm}|j\rangle, \quad (1.22)$$

then the elements of  $\rho$  are

$$\rho_{jk} = \langle j|\rho|k\rangle = \sum_m p_m c_{jm} c_{km}^*. \quad (1.23)$$

So the  $j^{\text{th}}$  diagonal element of  $\rho$  is

$$\rho_{jj} = \langle j|\rho|j\rangle = \sum_m p_m |c_{jm}|^2. \quad (1.24)$$

Since  $|c_{jm}|^2$  is the (conditional) probability of finding the system in the state  $|j\rangle$  given that it is initially in the state  $|\phi_m\rangle$ ,  $\rho_{jj}$  is the total probability of finding the system in the state  $|j\rangle$ .

If the density matrix consists of only a single state, so that  $\rho = |\psi\rangle\langle\psi|$ , then it is described as being *pure*. If it consists of a sum over more than one state, then it is described as being *mixed*, and the system is said to be in a *statistical mixture*, or simply *mixture*, of states. As an example, if the system is in the pure state

$$\rho = |\psi\rangle\langle\psi| = \sum_{jk} c_{jm} c_{km}^* |j\rangle\langle k| \quad (1.25)$$

is it said to be in a superposition of the basis states  $|j\rangle$ . If it is in the state

$$\rho = \sum_j p_j |j\rangle\langle j|, \quad (1.26)$$

then it is said to be in a mixture of the basis states  $|j\rangle$ . In the latter case, we can think of the system as really being in one of the basis states, and the density matrix merely describes the fact that we do not know which one (although some may be more likely than others). On the other hand, if the system is in a superposition of two states, then we cannot describe the system as really being in one state or the other, and we discuss this further in the next section.

The density matrix is said to be *completely mixed* if it is equal to  $I/N$ , where  $N$  is the dimensional of the system. If this is the case, then each of its eigenstates are equally likely. Further, *every* state in the Hilbert space is equally likely, since  $\langle\psi|(I/N)|\psi\rangle = 1/N$  for every state  $|\psi\rangle$ . In this case we have no information about the system.

It is not difficult to show that the density matrix is Hermitian ( $\rho = \rho^\dagger$ ), and has unit trace ( $\text{Tr}[\rho] = 1$ ). Since the density matrix is Hermitian it has a complete set of eigenvectors (eigenstates), and can always be diagonalized (that is, written in the

basis of its eigenstates so that it is diagonal). If the eigenvalues are  $\lambda_j, j = 1, \dots, N$ , and the corresponding eigenvectors are  $|\lambda_j\rangle$ , the density matrix is

$$\rho = \sum_j \lambda_j |\lambda_j\rangle \langle \lambda_j|. \quad (1.27)$$

So a system is never in a superposition of the eigenstates of the density matrix — it is either in a mixture of the eigenstates, or in a single eigenstate, in which case the density matrix is pure. Note that the eigenvalues are the probabilities of finding the system in the corresponding eigenstates.

### The difference between a superposition and a mixture

In case the physical distinction between a mixture of two or more quantum states, and a superposition of those states is not yet clear to you from your physics training to date, we pause here to explain this. A mixture of two states describes a situation in which a system really is in one of these two states, and we merely do not know which state it is. On the contrary, when a system is in a superposition of two states, it is *definitely not* in either of these states. The truth of this statement, while somewhat incredible, is easy to show. Consider the famous double-slit experiment, in which an electron is fired towards a metal sheet with two slits in it. The slits are very close to each other, and the electron is fired at the sheet so that it has a high probability of passing through the slits and thus reaching the other side. After the electron has passed through the slits, it is then detected on a screen some distance from the metal sheet. Consider what happens if one slit is blocked-up. The electron passes through the open slit, and is detected on the screen. If the state of the electron after passing through the slit and reaching the screen is  $|\psi_1\rangle$ , then the probability density that the electron is detected on the screen at position  $x$  is  $P_1(x) = |\langle \psi_1 | x \rangle|^2$ , where  $|x\rangle$  denotes the state in which the electron is on the screen at position  $x$ . Similarly, if the state of the electron after having passed through the other slit is  $|\psi_2\rangle$ , then the probability distribution for the electron to land on the screen at position  $x$  is  $P_2(x) = |\langle \psi_2 | x \rangle|^2$ .

Now we open both slits and fire the electron through them. If the electron *really* goes through one slit or the other, then we can immediately determine the probability distribution for the electron on the screen. Each electron that goes through the slits can be assigned as having gone through slit 1 or slit 2. Those going through slit one have probability distribution  $P_1(x)$ , and those going through slit 2 have distribution  $P_2(x)$ . If half of the electrons go through each slit (actually, some of the electrons will not pass through the slits, and instead scatter back from the metal sheet, but we ignore those) then the distribution of electrons on the screen will be

$$P(x) = \frac{1}{2}P_1(x) + \frac{1}{2}P_2(x). \quad (1.28)$$

This is the probability distribution we get if we assume that after going through the slits, the electron is in a *mixture* of states  $|\psi_1\rangle$  and  $|\psi_2\rangle$ . In this case the state of

the electron when it reaches the screen is  $\rho = (1/2)(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|)$ , and the probability distribution for being in state  $|x\rangle$  is

$$P_{\text{mix}}(x) = \text{Tr}[|x\rangle\langle x|\rho] = \frac{1}{2}P_1(x) + \frac{1}{2}P_1(x) = P(x). \quad (1.29)$$

However, after the electron passes through the slits, it is not in a mixture of states 1 and 2, but instead in the superposition  $|\psi\rangle = (|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}$ . Because of this the probability density for being in state  $|x\rangle$  is

$$\begin{aligned} P_{\text{sup}}(x) &= |\langle\psi|x\rangle|^2 = \frac{1}{2}P_1(x) + \frac{1}{2}P_1(x) + \text{Re}[\langle x|\psi_1\rangle\langle\psi_2|x\rangle] \\ &\neq P(x). \end{aligned} \quad (1.30)$$

Since  $P_{\text{sup}}(x)$  is not equal to  $P(x)$ , and  $P(x)$  is a necessary result of the electron *really* being in state  $|\psi_1\rangle$  or  $|\psi_2\rangle$ , a superposition cannot correspond to the electron really being in either of these two states. For want of better terminology, one usually refers to a system that is in a superposition of two states as being in *both states at once*, since this seems more reasonable than saying that it is in neither.

### A useful theorem: the most likely state

The following theorem answers the question, what is the most likely pure state of a system for a given density matrix  $\rho$ ? Note that the probability of finding the system in state  $|\psi\rangle$  is given by

$$P(|\psi\rangle) = \langle\psi|\rho|\psi\rangle = \text{Tr}[|\psi\rangle\langle\psi|\rho]. \quad (1.31)$$

**Theorem 1.** If the state of a quantum system is  $\rho$ , then the most likely pure state is the eigenstate of  $\rho$  with the largest eigenvalue. This eigenvalue is the probability that the system will be found in that state.

*Proof.* We need to find the state  $|\psi\rangle$  that maximizes  $\langle\psi|\rho|\psi\rangle$ . Let us denote the eigenbasis of  $\rho$  as  $\{|\lambda_n\rangle\}$ . Writing the density matrix in terms of its eigenbasis (Eq.(1.27)), and writing  $|\psi\rangle = \sum_n c_n|\lambda_n\rangle$ , we have

$$\langle\psi|\rho|\psi\rangle = \sum_n |c_n|^2 \lambda_n. \quad (1.32)$$

Since  $\sum_n |c_n|^2 = 1$ , the above expression is the average of the eigenvalues over the probability distribution given by  $p_n = |c_n|^2$ . Thus to maximize it we must place all of the probability on the largest eigenvalue. If we denote this eigenvalue by  $\lambda_j$ , then this means that  $|c_n|^2 = \delta_{nj}$ , and therefore  $|\psi\rangle = |\lambda_j\rangle$ .  $\square$

A density matrix consists of a set of states,  $\{|\psi_n\rangle\}$ , in which each state has an associated probability,  $p_n$ . Such a set of states and probabilities is called an *ensemble*. In fact, there is no reason why every state in an ensemble must be a pure state – a set

of states  $\{\rho_n\}$  with associated probabilities  $\{p_n\}$  is also an ensemble, and generates a density matrix  $\rho = \sum_n p_n \rho_n$ . If an ensemble contains only pure states, then it is called a *pure-state* ensemble. We will often write an ensemble for the set of states  $\{\rho_n\}$ , with associated probabilities  $\{p_n\}$ , as  $\{\rho_n, p_n\}$ .

Since every pure-state ensemble corresponds to some density matrix, a natural question to ask is, for a given density matrix, can one determine all the possible pure-state ensembles that correspond to it. It turns out that not only is the answer yes, but the collection of ensembles can be characterized very simply. We leave these facts to Chapter 2, however, because one of the characterizations uses the concept of *majorization*. This is a very simple concept which captures the intuitive notion of uncertainty, and will be defined in the next chapter.

### 1.2.3 Quantum measurements

In section 1.2.1 we introduced the measurement postulate, and defined the concept of a von Neumann measurement. Given a state-of-knowledge,  $\rho$ , we can describe a von Neumann measurement of the system in the following way. Given a basis  $\{|n\rangle\}$ , we define the set of projection operators

$$P_n \equiv |n\rangle\langle n|, \quad (1.33)$$

that project any initial state onto one of the basis states. After the measurement we will find the system to be in the state  $|m\rangle$  with probability

$$p_m = \langle m|\rho|m\rangle = \text{Tr}[P_m \rho P_m] = \text{Tr}[(P_m)^2 \rho], \quad (1.34)$$

where we have used the cyclic property of the trace:  $\text{Tr}[ABC] = \text{Tr}[CAB] = \text{Tr}[BCA]$ . The state after the measurement can be written as

$$\tilde{\rho}_m = |m\rangle\langle m| = \frac{P_m \rho P_m}{\text{Tr}[P_m \rho P_m]} = \frac{P_m \rho P_m}{p_m}. \quad (1.35)$$

We include the tilde on top of  $\rho_m$  to indicate that it is a state that results from a measurement. This convention will help to make our expressions clearer later on. The reason why we choose to write the expression for the final state using the projectors  $P_n$  is not, in fact, important at this point. It will, however, serve to make it clear in what follows that von Neumann measurements are a special case of the more general measurements that we will derive below.

Von Neumann measurements are certainly not the only kinds of measurements one can make on a quantum system, but fortunately we can derive all other measurements from them. To do this we consider a quantum system we wish to measure, which we will call the *target*, and a second system we will call the *probe*. The probe is prepared in some state independently of the target, and then the two system are allowed to interact. After the interaction we perform a von Neumann measurement on the probe. As a result of the interaction, the probe is correlated with the target,



and so the measurement on the probe provides us with information about the target. This procedure gives us the most general kind of measurement we can perform on a quantum system.

We will denote the basis in which we will measure the probe system as  $\{|n\rangle\}$ ,  $n = 0, \dots, N - 1$ , and since the interaction between the systems may be anything, we can start the probe in the state  $|0\rangle$  without any loss of generality. The initial state of the target and probe is therefore

$$\rho_{\text{comb}} = |0\rangle\langle 0| \otimes \rho, \quad (1.36)$$

where  $\rho$  is the initial state of the target. We will always write the state of the probe on the left hand side of the tensor product, and that of the target on the right. To understand the following it is essential to be familiar with how a composite quantum system, consisting of two subsystems, is described by the tensor product of the spaces of these two subsystems. If you are not familiar with how to combine two quantum systems using the tensor product, denoted by “ $\otimes$ ”, the details are presented in Appendix A.

### The form of the target/probe interaction operator

The interaction between the target and probe is described by some unitary operator  $U$ , that acts in the space of both systems. The subsequent von Neumann measurement on the probe is described by a projection onto one of the probe states  $|n\rangle$ , followed by a normalization. In analyzing the measurement process, it will make things clearest if we first know some things about the structure of  $U$ . (This is actually the hardest part conceptually — the rest is very straightforward. This section is, in fact, also quite easy to understand if you write  $U$  explicitly as a matrix containing sub-blocks, as you go along.)

Since  $U$  acts in the tensor-product space, it can be written as the matrix

$$U = \sum_{nn'kk'} u_{nk,n'k'} |n\rangle\langle n'| \otimes |s_k\rangle\langle s'_k|, \quad (1.37)$$

where  $|s_k\rangle$  are a set of basis states for the target, and  $u_{nk,n'k'}$  are the matrix elements of  $U$ . For each pair of probe states  $|n\rangle$  and  $|n'\rangle$ , there is a sub-block of  $U$  that acts in the space of the target. We can alternatively write  $U$  in terms of these sub-blocks as

$$U = \sum_{nn'} A_{nn'} \otimes |n\rangle\langle n'|, \quad (1.38)$$

where the operators  $A_{nn'}$  are given by

$$A_{nn'} = \sum_{kk'} u_{nk,n'k'} |s_k\rangle\langle s'_k|. \quad (1.39)$$

We will also denote the sub-blocks  $A_{n0}$  simply as  $A_n$ . These constitute the first column of sub-blocks of  $U$ . Because  $U$  is unitary,  $U^\dagger U = I$ . This means that each

of the diagonal sub-block of  $U^\dagger U$  that act on the target must be the identity operator on the target space. Considering the diagonal sub-block corresponding to the probe matrix element  $|0\rangle\langle 0|$ , and using Eq.(1.38), this means that

$$I = \langle 0|UU^\dagger|0\rangle = \sum_n A_n^\dagger A_n. \quad (1.40)$$

The final important fact about the structure of  $U$  is that, apart from the restriction given by Eq.(1.40), we can choose the sub-blocks  $A_n \equiv A_{0n}$  to be *any* set of operators. This is because the relation given by Eq.(1.40) is alone sufficient to ensure that there exists a unitary  $U$  with the operators  $A_n$  as its first column of sub-blocks. To see why this is true, note that the operators  $A_n$ , being the first column of sub-blocks, constitute the first  $M$  columns of  $U$ , where  $M$  is the dimension of the target. To show that Eq.(1.40) ensures that  $U$  is unitary, we need merely show that Eq.(1.40) implies that these first  $M$  columns are mutually orthonormal. So long as this is the case, we can always choose the remaining columns of  $U$  so that all the columns of  $U$  form an orthonormal basis, making  $U$  unitary. This task is not especially difficult, and we leave it as an exercise.

### Using a probe system to make a measurement

We now apply the unitary interaction  $U$  to the initial state of the two systems, and then project the probe system onto state  $|n\rangle$ . The final state of the combined system is then

$$\check{\sigma} = (|n\rangle\langle n| \otimes I)U(|0\rangle\langle 0| \otimes \rho)U^\dagger(|n\rangle\langle n| \otimes I). \quad (1.41)$$

We have placed a caron ( $\check{\cdot}$ ) over the final state,  $\sigma$ , to indicate that it is not necessarily normalized, due to fact that we have projected the system onto a subspace. Writing  $U$  in terms of its sub-blocks, we immediately obtain a simple form for the final state:

$$\check{\sigma} = |n\rangle\langle n| \otimes A_n \rho A_n^\dagger. \quad (1.42)$$

The final state of the *target* is therefore

$$\tilde{\rho}_n = \frac{A_n \rho A_n^\dagger}{\text{Tr}[A_n^\dagger A_n \rho]}, \quad (1.43)$$

where we have normalized it by dividing by its trace.

Now that we have a simple form for the final state for each outcome  $n$ , we need to know the probability,  $p_n$ , that we get each of these outcomes. This is given by the probability of finding the probe in state  $|n\rangle$  after the interaction  $U$ , and this is in turn the sum of all the diagonal elements of the density matrix that correspond to the probe being in state  $|n\rangle$ . The density matrix after the interaction is given by  $\rho_U = U(|0\rangle\langle 0| \otimes \rho)U^\dagger$ . The sum of its diagonal elements that correspond to the

probe state  $|n\rangle$  is therefore

$$\begin{aligned}
p_n &= \text{Tr} [ (|n\rangle\langle n| \otimes I) \rho_U (|n\rangle\langle n| \otimes I) ] \\
&= \text{Tr} [ (|n\rangle\langle n| \otimes I) U (|0\rangle\langle 0| \otimes \rho) U^\dagger (|n\rangle\langle n| \otimes I) ] \\
&= \text{Tr}[\check{\sigma}] \\
&= \text{Tr}[|n\rangle\langle n| \otimes A_n \rho A_n^\dagger] \\
&= \text{Tr}[|n\rangle\langle n|] \text{Tr}[A_n \rho A_n^\dagger] \\
&= \text{Tr}[A_n^\dagger A_n \rho].
\end{aligned} \tag{1.44}$$

We now have a complete description of what happens to a quantum system under a general measurement process. Further, we know that *every* set of operators  $\{A_n\}$  that satisfies Eq.(1.40) describes a measurement that can be realized by using an interaction with a probe system. For emphasis we state this result as a theorem.

**Theorem 2. The fundamental theorem of quantum measurement:** Every set of operators  $\{A_n\}$ ,  $n = 1, \dots, N$ , that satisfy  $\sum_n A_n^\dagger A_n = I$ , describes a possible measurement on a quantum system via the relations

$$\tilde{\rho}_n = \frac{A_n \rho A_n^\dagger}{p_n}, \tag{1.45}$$

$$p_n = \text{Tr}[A_n^\dagger A_n \rho], \tag{1.46}$$

where  $\rho$  is the state of the system before the measurement,  $\tilde{\rho}_n$  is the state of the system upon obtaining measurement result  $n$ , and  $p_n$  is the probability of obtaining result  $n$ .

For a given measurement, we will refer to the corresponding operators  $\{A_n\}$  as the *measurement operators* for the measurement. Now we have a simple mathematical description for general quantum measurements, but we don't yet know the physical meaning of the operators  $\{A_n\}$ . That is, we don't know how the form of a particular operator  $A_n$  corresponds to the physical properties of the measurement. We will explain this in section 1.3, but before we do, we discuss a couple of related topics.

### Discarding a subsystem

Let us say we have a quantum system  $\mathcal{S}$  that is initially in a state  $\rho$ . It then interacts with a second system  $\mathcal{B}$  over which we have no control, and to which we have no access. What is our state-of-knowledge of the first system after the interaction? One way to answer this question is to realize that another observer could measure the second system, just as we measured the probe in the scenario above. Since we do not know the outcome of the measurement, our state-of-knowledge is given by averaging

over all the outcomes. It is therefore

$$\begin{aligned}
 \tilde{\rho} &= \sum_n p_n \tilde{\rho}_n = \sum_n A_n \rho A_n^\dagger \\
 &= \sum_n \langle n | U(|0\rangle\langle 0| \otimes \rho) U^\dagger |n\rangle \\
 &\equiv \text{Tr}_B[U(|0\rangle\langle 0| \otimes \rho) U^\dagger]
 \end{aligned} \tag{1.47}$$

So our state-of-knowledge after the interaction is given by the partial trace over the second system. In fact, the partial trace is completely independent of the measurement that the observer makes on the second system. Of course, this is crucial, because we cannot know what measurement the observer will make, so our state-of-knowledge must be independent of this measurement. The definition of the partial trace, and its invariance under operations on the traced system, are described in Appendix A.

### Positive-Operator Valued Measures (POVM's)

General measurements on quantum systems are often referred to in the literature as “positive-operator valued measures”. While the usefulness of this term is unclear, and should probably be replaced simply with “quantum measurement”, its origin is as follows. A *measure* is a map that associates a number with all subsets of a given set. A positive-operator valued measure is thus a map that associates a positive operator with every subset. For the purposes of quantum measurements, the set in question is the set of outcomes of the measurement. Let us label the outcomes by  $n$ . If we pick some subset of the outcomes, call it  $\mathcal{M}$ , then the probability that we get an outcome from this subset is given by

$$\text{Prob}(n \in \mathcal{M}) = \sum_{n \in \mathcal{M}} p_n = \sum_{n \in \mathcal{M}} \text{Tr}[A_n^\dagger A_n \rho] = \text{Tr} \left[ \left( \sum_{n \in \mathcal{M}} A_n^\dagger A_n \right) \rho \right]. \tag{1.48}$$

Since the operators  $A_n^\dagger A_n$  are all positive, the operator in the parentheses is also positive. This is the positive operator associated with the set  $\mathcal{M}$ , and is thus the positive operator of the “positive-operator valued measure”.

## 1.3 Understanding Quantum Measurements

### 1.3.1 Relationship to classical measurements

While we have shown that a general measurement on a quantum system is described by a set of operators  $\{A_n\}$ , we do not yet understand how the form of these operators relates to the nature of a given measurement. We can gain such an understanding by examining various key forms that the  $A_n$  can take, and the measurements to which

they correspond. It is most informative to begin by examining under what conditions quantum measurements reduce to classical measurements.

Consider a state-of-knowledge that is diagonal in the basis  $\{|n\rangle\}$ , so that

$$\rho = \sum_n p_n |n\rangle\langle n|. \quad (1.49)$$

Note that in this case the system is not in a superposition of the basis states. It actually *is* in one of the basis states, we merely do not know which one. Recall that we refer to this situation by saying that the system is in a *statistical mixture* (or simply a *mixture*) of the states  $\{|n\rangle\}$ . Now, if we make a measurement on the system purely to obtain information about the basis state it is in, then we might expect this to be described perfectly well by a classical measurement. This is because the basis states are distinguishable, just like classical states, and our state-of-knowledge is a probability distribution over these states, just like a classical state of knowledge.

With the above thoughts in mind, consider a quantum measurement in which all the measurement operators  $A_j$  are diagonal in the same basis as the density matrix. Note that we can think of the diagonal elements of the measurement operators, and of  $\rho$ , simply as functions of the index  $n$ . That is, we can write

$$\rho = \sum_n P(n) |n\rangle\langle n|, \quad (1.50)$$

$$A_j = \sum_n A(j, n) |n\rangle\langle n|, \quad (1.51)$$

where  $P(n)$  and  $A(j, n)$  are the functions in question. The final (or posterior) state-of-knowledge that results from our measurement is

$$\tilde{\rho}_j = \frac{A_j \rho A_j^\dagger}{\text{Tr}[A_j^\dagger A_j \rho]} = \frac{\sum_n A^2(j, n) P(n) |n\rangle\langle n|}{\mathcal{N}}. \quad (1.52)$$

This is precisely the same form as the update rule for classical measurement theory. All we have to do is write the posterior state  $\tilde{\rho}_j$  as

$$\tilde{\rho}_j = \sum_n P(n|j) |n\rangle\langle n|, \quad (1.53)$$

and identify  $A^2(j, n)$  as the likelihood function,  $P(j|n)$ , and we have

$$P(n|j) = \frac{P(j|n)P(n)}{\mathcal{N}}. \quad (1.54)$$

So, at least in this case, quantum measurements are merely classical measurements, and the diagonal of the measurement operator  $A_j$  corresponds to the square-root of the likelihood function evaluated for the  $j^{\text{th}}$  outcome.

Two operators  $A$  and  $B$  are diagonal in the same basis if and only if they commute with one another. In view of the above results, we will refer to quantum measurements for which

$$[A_j, A_k] = [A_j, \rho] = 0, \quad \forall j, k \quad (1.55)$$

as *semi-classical measurements*, since they are completely equivalent to their classical counterparts. (We would prefer to call them simply *classical measurements*, but refrain from this to avoid confusion.)

We will see in section 1.3.3 that we can, in fact, understand the action of all quantum measurements, to a large extent, in terms of Bayesian inference.

### 1.3.2 Incomplete measurements and Bayesian inference

As an example we consider a simple semi-classical measurement that provides only partial information about the final state. Such measurements have measurement operators that are not projectors onto a single state, and are referred to variously as “incomplete” measurements [11], “finite-strength” measurements [12], or “weak” measurements [13].

Consider a two-state system in the state

$$\rho = p_0|0\rangle\langle 0| + p_1|1\rangle\langle 1|, \quad (1.56)$$

and a measurement with two possible outcomes, described by the operators

$$A_0 = \sqrt{\kappa}|0\rangle\langle 0| + \sqrt{1-\kappa}|1\rangle\langle 1|, \quad (1.57)$$

$$A_1 = \sqrt{1-\kappa}|0\rangle\langle 0| + \sqrt{\kappa}|1\rangle\langle 1|. \quad (1.58)$$

If  $\kappa = 1$  (or  $\kappa = 0$ ) then both measurement operators are projectors, and after the measurement we are therefore certain about the state of the system. If  $\kappa = 1/2$  then both operators are the identity. They therefore have no action on the system, and we learn nothing from the measurement. For  $\kappa \in (1/2, 1)$  the measurement is an “incomplete” measurement. It changes the probabilities  $p_0$  and  $p_1$ , but does not reduce either to zero.

As discussed in the previous section, since the measurement operators commute with the density matrix, their action on the diagonal elements of the density matrix is simply that of Bayesian inference. The measurement operators we have chosen above correspond precisely to the simple two-outcome measurement that we discussed in section 1.1.1, with the identification  $\alpha = \kappa$ . So for example, if  $\kappa > 1/2$  and we get the outcome 0, then the diagonal elements of  $\rho$  are re-weighted so as to make the state  $|0\rangle$  more likely than it was before. Finite-strength measurements thus provide some information about which state the system is in, but not complete information.

### 1.3.3 The polar decomposition: bare measurements and feedback

The polar decomposition theorem states that any operator  $A$  may be written as the product of a positive operator,  $\mathcal{P}$ , and a unitary operator  $U$  [14, 15]. (The unitary operator may be on the right or the left of the positive operator.) Positive operators have a complete set of eigenvalues, and all their eigenvalues are greater than or equal to zero (the density matrix is thus a positive operator). As a result  $\mathcal{P} = \mathcal{P}^\dagger$ . When a unitary operator,  $U$ , is applied to the density matrix  $\rho$  via  $U\rho U^\dagger$ , it gives a possible time evolution for the system. In doing so it does not change the eigenvalues of the density matrix, but only the eigenvectors. This means that it does not change the amount of our uncertainty about a system, just *where* in the state space it is most likely to be. The eigenvalues of a unitary matrix are in general, complex but have unit norm, and  $U^\dagger U = U U^\dagger = I$ . As we will see, the polar decomposition theorem provides tremendous insight into the structure of quantum measurements.

To apply the polar decomposition theorem to a quantum measurement described by a set of operators  $\{A_n\}$ , we write each of these operators as  $A_n = U_n \mathcal{P}_n$ . With this replacement the probability of obtaining the  $n^{\text{th}}$  result becomes

$$p_n = \text{Tr}[(\mathcal{P}_n)^2 \rho], \quad (1.59)$$

and the final state can be written as

$$\tilde{\rho}_n = U_n \left( \frac{\mathcal{P}_n \rho \mathcal{P}_n}{p_n} \right) U_n^\dagger. \quad (1.60)$$

The requirement upon the operators  $A_n = U_n \mathcal{P}_n$  for them to represent a measurement becomes

$$\sum_n A_n^\dagger A_n = \sum_n (\mathcal{P}_n)^2 = I. \quad (1.61)$$

This representation of a quantum measurement has a number of important ramifications. First note that since the unitary operators  $U_n$  do not change the eigenvalues of the density matrix, they play no role in describing the information extraction by the measurement. The information gathering properties of the measurement are completely captured by the positive operators  $\mathcal{P}_n$ .

The unitary operators  $U_n$  are not constrained in any way. Note further that the form of Eq.(1.60) is exactly the same form we would obtain if we made a measurement described by the positive operators  $\{\mathcal{P}_n\}$ , and applied forces to the system to generate the evolution  $U_n$  upon getting the  $n^{\text{th}}$  result. Applying forces to a system conditional upon the result of a measurement is often referred to as *feedback*, and is the basis of the subject of feedback control.

The polar decomposition shows us that to understand the process of information gathering in any quantum measurement, we need only examine the action of positive operators on quantum states. We will see shortly that this allows us to understand much of the action of quantum measurements in terms of Bayesian inference. Before we examine positive operators in more detail, it is worth making a closer connection between the structures of quantum and classical measurement theory.

### The parallel structure of classical and quantum measurements

As introduced in section 1.1, classical measurement theory involves multiplying our initial probability distribution by the likelihood function (and normalizing) to obtain our final state-of-knowledge. The reason that we do not include any other action on the system is because classical measurements need not disturb the system at all. A classical system *really is* in one of its possible states  $n$ , and the only thing that changes during a measurement is our state-of-knowledge regarding the system. Of course, we could if we wanted include the effects of forces applied to the system dependent upon the result of the measurement.

In general the evolution of a classical system is given by a symplectic operator acting on a state in the system's phase-space. But for our purposes here we can think of our classical system as having a discrete set of states. A six-sided die, for example, where the state is the number that is face-up. We can represent the states of this system (the die) as the elementary basis vectors of a six-dimensional vector space. With this representation, the difference between a quantum and classical system is that the former can be in any superposition of the basis states, whereas the latter can only be in one of the basis states. With this representation for a classical system we can use the formalism of quantum measurement theory: our state-of-knowledge is given by a density matrix that is diagonal in the basis defining the classical states, and all our measurement operators (the likelihood function) are also diagonal in this basis. With this representation, the only actions we can perform on our classical system are permutations of the classical states. Permutations of the basis vectors are given by applying permutation matrices. These are unitary matrices whose elements are either 0 or 1, and each row and column contains exactly one 1.

Defining  $\{T_n\}$  to be a set of permutation matrices, then a classical measurement followed by a deterministic action on the system is given by

$$\tilde{\rho}_n = T_n \left( \frac{\mathcal{P}_n \rho \mathcal{P}_n}{p_n} \right) T_n^\dagger. \quad (1.62)$$

where  $[\mathcal{P}_n, \mathcal{P}_m] = 0, \forall n, m$ . This exactly parallels a quantum measurement. This suggests that for quantum measurements one can view the positive operators as taking the place of the likelihood functions of classical measurement theory. The unitary part of the operators that describe quantum measurements play the role of the permutation matrices (or symplectic evolutions) in the classical theory (and for this reason are not usually included in the classical theory). When we consider the subject of information in the next chapter, we will see that there are further reasons to view the positive operators as describing the quantum analogue of classical measurements.

### The action of bare measurements

Since it is the positive operators that characterize the information extraction in a quantum measurement process, we will refer to quantum measurements that are described purely by a set of positive operators,  $\{\mathcal{P}_n\}$ , as *bare* measurements. This



name is intended to indicate that there is a sense in which they capture the notion of information extraction without additional actions<sup>1</sup>.

Now let us examine the action of bare measurements (positive operators) on a quantum system. First recall that while a semi-classical measurement changes our state-of-knowledge of a quantum system, it need not change the underlying state-vector. To see this, consider a semi-classical von Neumann measurement. This is a von Neumann measurement that projects the system onto the basis in which the density matrix is diagonal. If the density matrix is diagonal in the basis  $\{|n\rangle\}$ , then the system could really be in one of the basis states. If this is the case, the action of the von Neumann measurement is merely to select the basis state that the system is already in. In this case it does not change the state of the system, merely our information about it.

Contrary to the behavior of classical measurements, however, von Neumann measurements, and thus bare measurements in general, *can* change the state of a system. If we prepare a system in a pure state that is *not* one of the basis states onto which a von Neumann measurement projects, then the measurement necessarily changes the state of the system. More generally, a positive operator will change the state of a system whenever the operator does not commute with the density matrix. Nevertheless, the action of a positive operator on a quantum system can be understood in quite simple terms. To do this we go to the basis in which the positive operator,  $\mathcal{P}$  is diagonal. Let us call this the  $\mathcal{P}$  basis. In general the density operator will not be diagonal in this basis. Recall that even so, the diagonal elements of  $\rho$  do still give the probability distribution over the  $\mathcal{P}$  basis states. Now, if we examine the action of  $\mathcal{P}$  on the diagonal elements of  $\rho$  in the  $\mathcal{P}$  basis, we find that this is exactly the same it would be if  $\rho$  were diagonal in this basis. That is, as far as the measurement is concerned, the action of  $\mathcal{P}$  on  $\rho$  is precisely to apply Bayes' theorem to the distribution over the  $\mathcal{P}$  basis. (In doing so the action of  $\mathcal{P}$  also changes the off-diagonal elements of  $\rho$ , re-weighting them in a manner which is consistent with the way it re-weights the diagonal elements.)

We see from the above discussion that we can think of each of the positive operators,  $\mathcal{P}_n$ , of a bare measurement as extracting information about the basis in which they are diagonal (their eigenbases). The difference between a bare quantum measurement and a classical measurement is that the operators of a bare measurement may each extract information about a *different* basis.

### Summary

The polar decomposition theorem shows that the measurement operators describing a given measurement can be factored into the product of a positive operator and a

---

<sup>1</sup>We note that “bare” measurements have also been referred to in the literature as “minimally disturbing” measurements [16, 17]. While these two terms are similar in spirit, we prefer the latter for two reasons. The first is that whether or not bare measurements are minimally disturbing does depend on ones definition of disturbance. (Three useful notions of disturbance are discussed in section 2.3.2.) The second is its brevity.

unitary operator. It is the positive operators that describe the information extraction. Since the unitary operators describe the application of forces to a system, and since in general they will depend upon the outcome of the measurement, they can be interpreted as a feedback process. Consequently we can think of the polar decomposition theorem as providing an interpretation of all measurements, and thus all quantum evolutions, as a combination of information extraction and feedback.

Since the action of positive operators are identical to classical measurements when the density operator is diagonal in their eigenbasis, one can regard them as describing information extraction without additional actions. That is, the effect on the quantum state due to a positive measurement operator can be regarded, if one wishes, as only that which is a direct result of information extraction, and nothing more.

## 1.4 Describing Measurements within Unitary Evolution

Measurement does introduce an extra feature into quantum mechanics, over and above the unitary (and thus reversible) evolution given by Schrödinger's equation. However, it is very important to realize that one can treat measurements, as well as the actions that an observer performs on a system as a result of those measurements, entirely using unitary evolution (well, *almost* entirely). This is important for at least two reasons. It provides a simple way to understand how measurement processes fit into thermodynamics and the second law (Chapter 4), and it allows us to make a direct connection between different ways of implementing feedback control in quantum systems (Chapter 6).

Once an observer has made a measurement, she can apply forces to the system that depend on the result of the measurement. So to describe measurement we need to be able to describe an arbitrary sequence of measurements, where each measurement is followed by a unitary operation on the system that depends upon (is “conditional on”) the measurement result. Our analysis of measurements in terms of a probe system in Sec. 1.2.3 already provides the tools to do this using unitary evolution, as we now show.

Consider making a general measurement with a probe system. The result of the measurement is given by projecting the probe system onto one of a set of orthogonal states. This means that we can determine the result of the measurement just by looking at what has happened in each of a set of mutually orthogonal subspaces that correspond to the mutually orthogonal states of the probe. To apply a unitary operation based on each measurement result, we do not actually need to project the probe system onto these orthogonal subspaces. Instead, we could design the interaction Hamiltonian so that its action on the system depends on the subspace of the probe. In fact, this conditional action is already incorporated into the interaction operator  $U$  that creates the measurement: by choosing  $U$  appropriately, we can choose each measurement operator  $A_n$ , and in particular we can choose the  $A_n$  to include the desired conditional unitary action on the system. Thus our construction of a measurement using a probe already contains both measurement and actions performed

as a result of that measurement, without performing any projection of the probe. Note that there is also no need for the subspaces be one-dimensional: we can always choose  $U$  so that each measurement result corresponds to an arbitrarily large subspace of the probe.

If we want to make a sequence of measurements, along actions based on the result of each measurement, then we could use a sequence of probe systems, each one interacting with the system. Alternatively, we could swap the state of the probe system with a third “auxiliary” system, prepared in state  $|0\rangle$ . This can be done with a unitary transformation acting in the space of the probe and the third system. It prepares the probe in state  $|0\rangle$ , and stores the measurement result (the subspaces corresponding to the different outcomes of the measurement) in the third system. The probe is then ready to make another measurement on the system. In this way multiple measurements can be made on the system, the results of each measurement being stored in a new auxiliary system. The whole process is purely unitary evolution in the space of the system, the probe, and the auxiliary systems. Note that we could also make the choice of subsequent measurements conditional on the results of previous measurements. We can do this by having the subsequent measurements implemented by unitary interactions that use only the subspace of the probe that corresponds to previous measurement results.

The only time we need to invoke an explicit projection (a “real” measurement) is if we want to specify that only one sequence of measurement results “actually” occurred. In that case, we perform one final projection onto the subspaces of the set of auxiliary systems, and that gives us the final state of the system that results from that specific sequence of results. But note that the unitary procedure actually contains *all* the information about the behavior of the system under every possible sequence of measurement results. This information is contained within a set of orthogonal subspaces, defined by the (mutually orthogonal) states of the auxiliary systems.

A special case of the above procedure is when the conditional actions on the system are designed so as to prepare the system in the same final state, regardless of the outcomes of the measurement. This is possible because the conditional unitary evolutions can map the states that result from each measurement, to a predetermined final state, so long as all these states are pure (or if their density matrices all have the same set of eigenvalues). In this case the final state of the system does not depend on the subspaces of the probe or auxiliary systems, and is thus fully determined without any projections. Thus in this case the entire sequence of measurements and conditional operations is fully described without projections.

In the procedure described above for realizing a sequence of generalized measurements, we assumed that we could turn off and on the required unitary evolutions between the various subsystems. This would appear to require the action of an additional external system. In fact, the use of unitary evolutions that are switched off and on was purely a conceptual tool. No such external system is required, because the whole sequence of measurements can be performed with a single unitary, generated by a single time-independent Hamiltonian. This can be seen by noting that

the sequence of unitary evolutions that are switched off and on, produce at the final time one unitary map from the initial to final state of the whole system. Thus the whole evolution can be described by a single unitary generated by a single Hamiltonian. Another way to see that the whole process can be performed with a single unitary is that a sequence of generalized measurements is equivalent to a single generalized measurement, where the latter has a single outcome corresponding to every possible sequence of measurement results. This single generalized measurement can therefore be implemented with a single probe and a single unitary  $U$ .

## 1.5 Inefficient Measurements

Inefficient measurements are defined as measurements in which the observer does not have access to all the information regarding the measurement results. We can describe this situation by labeling the measurement operators with two indices:  $\{A_{jk}\}$ . We give the observer access to the value of  $j$ , but not the value of  $k$ . Thus the observers final state-of-knowldge is given by averaging over the index  $k$ , and is

$$\tilde{\rho} = \sum_k P(k|j) \tilde{\rho}_{jk} = \sum_k \frac{P(j,k)}{P(j)} \tilde{\rho}_{jk} = \frac{1}{P(j)} \sum_k A_{jk} \rho A_{jk}^\dagger. \quad (1.63)$$

where

$$\tilde{\rho}_{jk} = \frac{A_{jk} \rho A_{jk}^\dagger}{P(j,k)}, \quad (1.64)$$

and  $P(j,k) = \text{Tr}[A_{jk}^\dagger A_{jk} \rho]$ .

Of course, the most general situation is when we have just a single index for the measurement result,  $n$ , and the observer has some general classical state-of-knowledge regarding this result. We can describe this situation by saying that the observer makes a classical measurement of  $n$ , and get result  $m$ , where this classical measurement is described by the likelihood function  $P(m|n)$ . The formulation above, while appearing to be simpler, actually covers this most general scenario. To see this note that in this general case the observers final state state-of-knowledge is given by

$$\begin{aligned} \tilde{\rho} &= \sum_n P(n|m) \tilde{\rho}_n = \sum_n \frac{P(m|n)P(n)}{P(m)} \tilde{\rho}_n \\ &= \frac{1}{P(m)} \sum_n P(m|n) B_n \rho B_n^\dagger, \end{aligned} \quad (1.65)$$

where the measurement operators are  $\{B_n\}$ . Now we observe that we can obtain this result using the two-index formulation above by setting  $m = j$ ,  $n = k$ , and

$$A_{mn} = \sqrt{P(m|n)} B_n. \quad (1.66)$$

A probe system can be used to implement inefficient measurements by choosing the initial state of the probe to be mixed. Alternatively one can simply introduce a randomization process (a classical noise process) between the measurement result and the data actually obtained by the observer.

## 1.6 Measurements on Ensembles of States

It is sometimes useful to know not only the density matrix for a system, but also the full set of possible pure states that the system might be in, along with their respective probabilities. This may be important if we know that someone has prepared the system in one of a specific set of states, and we wish to make a measurement to determine which one. A set of states  $\{|\psi_j\rangle\}$ , along with their corresponding probabilities,  $\{P(j)\}$ , is referred to as an *ensemble*, and we will often write it as  $\{|\psi_j\rangle P(j)\}$ . Of course, there is no reason why an ensemble need contain only pure states, and so in general we will write our ensembles as  $\{\rho_j, P(j)\}$ . So far we have examined how the action of a measurement changes  $\rho$ , but it is also useful to know how it changes an ensemble.

If the initial ensemble is  $\{\rho_j, P(j)\}$ , then the density matrix is

$$\rho = \sum_j P(j) \rho_j. \quad (1.67)$$

If we now make a measurement described by the measurement operators  $\{A_n\}$ , and obtain the outcome  $n$ , then the final state is

$$\begin{aligned} \tilde{\rho}_n &= \frac{A_n \sum_j P(j) \rho_j A_n^\dagger}{\mathcal{N}} \\ &= \frac{1}{\mathcal{N}} \sum_j P(j) \text{Tr}[A_n^\dagger A_n \rho_j] \left( \frac{A_n \rho_j A_n^\dagger}{\text{Tr}[A_n^\dagger A_n \rho_j]} \right) \end{aligned} \quad (1.68)$$

Now let us examine the various elements in the above expression. The quantum state in the final parentheses is correctly normalized. It is the state the system would be in following the measurement if the system had initially been in the state  $|\psi_n\rangle$ . We shall therefore denote it by  $\tilde{\rho}_{j,n}$ , meaning the state of the system given the values of both  $n$  and  $j$ .

The quantity  $\text{Tr}[A_n^\dagger A_n \rho_j]$  is the probability that we obtain result  $n$ , if the system is initially in the  $j^{\text{th}}$  state, so we can write this as the likelihood function  $P(n|j)$ . With these definitions the final state is now

$$\tilde{\rho}_n = \sum_j \left[ \frac{P(n|j)P(j)}{\mathcal{N}} \right] \tilde{\rho}_{j,n}. \quad (1.69)$$

We can also ask, what does the measurement tell us about the state that the system was in *before* the measurement. In classical measurement theory these two questions

are the same, since classical measurements do not change the state of the system. But for quantum systems the two questions are distinct. Because we know the initial ensemble, we know that the system was initially in one of the states  $\{\psi_n\}$ . To determine what the measurement tells us about the initial preparation all we have to do is to apply Bayesian inference to our classical state of knowledge regarding which of the possible states was prepared, being  $P(n)$ . After the measurement our state-of-knowledge regarding which state was prepared is thus

$$P(j|n) = \frac{P(n|j)P(j)}{\mathcal{N}}, \quad (1.70)$$

and this is also the expression appearing in Eq.(1.69). So we can also write the state of the system after the measurement as

$$\tilde{\rho}_n = \sum_j P(j|n)\tilde{\rho}_{j,n}, \quad (1.71)$$

which, if you think about it, makes perfect sense.

### Summary

An important message of the above discussion is that when we make a measurement on a quantum system, we can consider two distinct kinds of information that the measurement extracts. The first is information about the *final* state of the system – this concerns how certain we are of the state of if the system following the measurement. This is the kind of information that is useful when we are trying to control a system. The second kind of information is information about the *initial* state of the system. As we will see in the next chapter (specifically, section 2.3), this is useful in a scenario in which the quantum system is being used to *communicate* information from one person to another. If Alice wishes to send a message to Bob, then she prepares the system in one of set of states, where each state represents a possible message, and sends the system to Bob. Bob then makes a measurement on the system to determine the initial state, and hence the message. In the next chapter we will learn how both kinds of information can be quantified.

### History and Further Reading

Bayes' theorem, the cornerstone of classical measurement theory, was introduced by Bayes in 1763 [18]. Further details regarding Bayesian inference and methods for finding priors are given in Jaynes' book on probability theory [4]. Many examples of the application of Bayesian inference can be found in [19]. The formalism describing general quantum measurements was first developed by Krauss [20]. The polar decomposition theorem can be found in "Matrix Analysis" by Horn and Johnson [14], and "Matrix Analysis" by Rajendra Bhatia [15].

The problem of finding the optimal measurement to distinguish between the states of a given ensemble appears to have no analytic solution. A number of properties of the optimal measurement are known, however, including approximately optimal measurements. Conditions that the optimal measurement must satisfy were first obtained by Holevo [21] and Yuen *et al.* [22]. Further details regarding this question may be found in [23, 24]. The latter gives an overview of the approximately optimal measurements that are known to date.

## Exercises

1. Show that all the eigenvalues of the density matrix are non-negative.
2. Show that  $\text{Tr}[\rho^2]$  depends only on the eigenvalues of  $\rho$ .
3. Consider a state-of-knowledge (density matrix) in which every pure-state is equally likely. Show that in this case the density matrix is proportional the identity.
4. The polar decomposition theorem states that every operator  $A$  can be written as  $UP$ , or  $QV$ , where  $P$  and  $Q$  are positive, and  $U$  and  $V$  are unitary. Show that  $P = \sqrt{A^\dagger A}$  and  $Q = \sqrt{AA^\dagger}$ .
5. Show that if a matrix  $A$  has a complete set of orthogonal eigenvectors (that is,  $A$  is diagonalizable) then  $A^\dagger A = AA^\dagger$ . (Thus, operators that do not satisfy this condition cannot be diagonalized.) In fact, the converse is also true — every matrix  $A$  for which  $A^\dagger A = AA^\dagger$  has a complete set of orthogonal eigenvectors [14, 25].
6. Consider a measurement that has two outcomes. Show that in this case the positive operators associated with each of the measurement operators commute with one another.
7. Consider the unitary operator  $U$  defined in Eq.(1.38). Show that Eq.(1.40) implies that the first  $M$  columns of  $U$  are orthonormal.
8. Consider a two-state system with basis states  $|0\rangle$  and  $|1\rangle$ . The system is prepared in the state  $|\psi_1\rangle = |1\rangle$ , or the state  $|\psi_2\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ , with equal probability. We now make a two-outcome measurement on the system to determine which of the two states it is in. We choose one of the measurement operators to be the projector  $P_0 = |0\rangle\langle 0|$ . This ensures that, if this outcome occurs, we can be *certain* that the initial state was  $|\psi_2\rangle$  (Why?). Determine the probability that, following the measurement, we will be uncertain of the initial preparation.
9. Show that any two measurements, performed one after the other, is equivalent to a single measurement, and determine the measurement operators for the

single measurement in terms of the first two. The result clearly generalizes for a sequence of many measurements.

10. Consider an efficient measurement performed on a system in a pure state. Show that by using feedback (that is, by applying to the system a unitary operator that depends upon the measurement result), one can prepare the system in the same final state for every outcome of the measurement, and this can be done for *any* efficient measurement. Is the same true if the initial state is mixed? If not, why not?
11. Show that if a system is prepared in one of two states that are not orthogonal to one another, there is no measurement that can guarantee to determine with certainty which state the system was prepared in. (You can use the construction involving a probe system, and the fact that unitary evolution does not change the inner product between two states.)
12. Consider a system prepared in one of two non-orthogonal states, where the probability of each is  $1/2$ . In this case we can describe the system entirely using a two-dimensional state-space (why?). Denoting our basis-states as  $|0\rangle$  and  $|1\rangle$ , we can write the two states as  $|\pm\rangle = \alpha|0\rangle \pm e^{i\phi}\sqrt{1-\alpha}|1\rangle$ . To determine which measurement will most accurately determine which of these two states the system is in, we must quantify what we mean by “most accurate”. In fact, different goals will motivate different definitions, and thus different optimal measurements. One such optimal measurement, first determined by Levitin [26], is the following. It has two outcomes, and the corresponding measurement operators,  $A_{\pm}$ , are, respectively, the projectors onto the two states  $|\pm\rangle_m = (|0\rangle \pm e^{i\phi}|1\rangle)/\sqrt{2}$ . Let say that if the outcome is  $+$ , then the observer guesses that the state was  $|+\rangle$ , and vice versa. i) Draw all the above states on the Bloch sphere. ii) Calculate the probability that the observer’s guess is incorrect (that the observer fails to correctly distinguish the states). (Note: this measurement maximizes the “mutual information”, to be discussed in Chapter 2, and does not maximize the probability of a successful guess. Can you think of how the measurement could be altered to increase the probability that the observers guess is correct?)
13. A system is prepared in one of the two orthogonal states  $|0\rangle$  and  $|1\rangle$ , each with probability  $1/2$ . A measurement is made on the system which has two outcomes given by the operators  $A_{\pm}$ . These operators project the system onto the states  $|\pm\rangle = \alpha|0\rangle \pm \sqrt{1-\alpha}|1\rangle$ , respectively. i) Calculate the change in the observers state-of-knowledge when he gets each of the two outcomes. ii) Which value of  $\alpha$  provides the observer with no information about the initial state?
14. Find an example of an inefficient measurement that leaves the system in a completely mixed state, regardless of the initial state.



15. Consider a two-state system in state  $|0\rangle$  undergoing evolution given by the Hamiltonian  $H = (i/\hbar)\mu\sigma_y$ . The state of the system as a function of time is then  $|\psi(t)\rangle = \cos(\mu t)|0\rangle + \sin(\mu t)|1\rangle$ . An observer now makes a measurement that projects the system either onto state  $|0\rangle$  or state  $|1\rangle$ . Consider that the observer makes a measurement after a short time interval  $t = \Delta t \ll 1/\mu$ .
- i) What is the probability that the observer will project the system into state  $|1\rangle$ ? (Hint: expand the sine function as a power series in  $\Delta t$ , and drop all terms higher than second-order in  $\Delta t$ . This will be useful for the next part.)
- ii) Now consider what happens if the observer makes repeated measurements, interrupting the evolution at intervals  $\Delta t$  apart. What is the probability that the observer will find the system to be in state  $|1\rangle$  after  $N$  measurements? iii) Finally, consider an interval  $T$ , in which the observer makes  $N$  measurements, each separated by the interval  $\Delta T = T/N$ . Calculate the probability that the system is found to be in state  $|1\rangle$  for every measurement. Now determine this probability in the limit in which  $\Delta t \rightarrow 0$  ( $N \rightarrow \infty$ ). The result is that the system is never found to be in state  $|1\rangle$ : the evolution of the system is frozen by the measurement. This is the *quantum Zeno effect*.
16. The “anti”-quantum Zeno effect (it is best to do exercise 15 first): Consider a two-state system that starts in state  $|0\rangle$ . After a time  $\Delta t$  an observer makes a two-outcome measurement on the system, whose measurement operators project the system onto the states  $|+\rangle = \cos(\Delta t)|0\rangle + \sin(\Delta t)|1\rangle$ , and  $|-\rangle = \sin(\Delta t)|0\rangle - \cos(\Delta t)|1\rangle$ . Calculate the probability that the system is found to be in the state  $|-\rangle$  after the measurement. This result tells you that, to first-order in  $\Delta t$ , the system will *always* be found in state  $|+\rangle$  after the measurement. The measurement thus “drags” the system from state  $|0\rangle$  to state  $|+\rangle$ . This is called the anti-quantum-Zeno effect, and has precisely the same origin as the Zeno effect (see exercise 15 above).
17. Weak, or *incomplete* measurements are those that do not project an initially mixed state only onto pure states — that is, they are measurements whose measurement operators are not all rank 1, and thus do not provide the observer with full information about the state of the system following the measurement. Consider a two-state system that is initially in the state

$$\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|. \quad (1.72)$$

Now consider the two-outcome measurement with projection operators given by

$$A_+ = \sqrt{k}|+\rangle\langle +| + \sqrt{1-k}|-\rangle\langle -|, \quad (1.73)$$

$$A_- = \sqrt{1-k}|+\rangle\langle +| + \sqrt{k}|-\rangle\langle -|, \quad (1.74)$$

where

$$|+\rangle = \sqrt{\alpha}|0\rangle + \sqrt{1-\alpha}|1\rangle, \quad (1.75)$$

$$|-\rangle = \sqrt{1-\alpha}|0\rangle + \sqrt{\alpha}|1\rangle. \quad (1.76)$$

The purity of a quantum state  $\rho$  is defined as  $\text{Tr}[\rho^2]$ . (This measure of how pure the quantum state is — loosely how much information we have about the state of the system — will be discussed in Chapter 2.) i) Write a computer program in your favorite language (e.g. Matlab) to calculate the increase in the purity of the state due to the measurement, averaged over the two measurement outcomes. For fixed values of  $p$  and  $k$ , what value of  $\alpha$  gives the maximum average increase in the purity? ii) Draw the initial state  $\rho$  on the Bloch sphere, along with the states  $|+\rangle$  and  $|-\rangle$  for the measurement that gives the maximum increase in purity.