

# Probabilities

Robert B. Griffiths  
Version of 27 August 2012

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Sample Space, Event Algebra, Probabilities</b>	<b>2</b>
2.1	Classical . . . . .	2
2.2	Quantum . . . . .	3
<b>3</b>	<b>Random and Physical Variables</b>	<b>4</b>
3.1	Classical . . . . .	4
3.2	Quantum . . . . .	6
<b>4</b>	<b>Several Random (Physical) Variables</b>	<b>7</b>
4.1	Classical . . . . .	7
4.2	Quantum . . . . .	9
<b>5</b>	<b>Born Rule</b>	<b>10</b>
5.1	Pre-probabilities . . . . .	10
5.2	Measurements . . . . .	11

---

### References:

CQT = R. B. Griffiths, *Consistent Quantum Theory* (Cambridge, 2002)  
 Feller, *An introduction to probability theory and its applications*, Vol. 1, 3d ed (Wiley 1968).  
 DeGroot and Schervish, *Probability and Statistics*, 3d ed (Addison-Wesley, 2002)

## 1 Introduction

- Classical (ordinary) probability theory. This is treated in a large number of texts. Helpful treatments of the material needed for quantum mechanics will be found in:

Feller: Introduction, Ch. I, Ch. V

DeGroot and Schervish, Ch. 1 and Secs. 2.1, 2.2, 2.3, 3.1, 3.5, 4.1, 4.2.

- For the corresponding quantum formulation, see CQT Ch. 5. The material is not readily accessible elsewhere, and the treatment in quantum textbooks is inadequate.

★ Probabilities play an essential role in quantum dynamics, as was first pointed out in 1926 by Born.

- The rules for manipulating probabilities and forming probabilistic inferences are the same in quantum as in classical physics *provided* attention is paid to incompatible quantum properties. This requires that one be careful to observe certain rules which, while also present in the classical case, can be easily overlooked in the quantum situation, leading to problems and paradoxes.

★ The traditional approach in quantum textbooks is to employ *measurements* as a tool for introducing probabilities. This approach is not wrong, but it is inadequate. It suffices for formulating rules of calculation which give the right answer when used in the right way, but they have given

rise to a lot of confusion. This approach is also inconsistent in the sense that it does not provide a way of describing real measuring apparatus in quantum terms.

- One of the confusions produced by the traditional approach: shall we first sum (amplitudes) and then square (to get probabilities), or shall we first square and then sum?

- Here are some quotes from a modern textbook:<sup>1</sup>

“If the apparatus permits us in principle to determine which of the two slits a neutron passes through, the interference will be destroyed independently of whether we actually bother to determine which slit it was”

“In summary, we must sum the amplitudes for identical final states and the probabilities for different final states, even if these final states differ only by physical parameters other than those of interest. It is sufficient that these other parameters be accessible in principle, even if they are not actually observed, for us to consider the final states as being different.”

- If you already know what the author is trying to say, you can figure out what he means, but it is otherwise rather difficult. A better strategy is to employ a precise formulation of quantum principles, and then apply it to various examples.

## 2 Sample Space, Event Algebra, Probabilities

★ Probabilities, whether quantum or classical, are formulated in terms of three things: a *sample space*  $\mathcal{S}$ , an *event algebra*  $\mathcal{E}$ , and a probability distribution  $\mathcal{P}$ . One sometimes speaks of the triple  $(\mathcal{S}, \mathcal{E}, \mathcal{P})$ .

- References: CQT Sec. 5.1 is quite short. For more details see Feller, or DeGroot and Schervish, or other textbooks on probabilities

### 2.1 Classical

★ An ordinary or “classical” sample space  $\mathcal{S}$  consists of a set of *mutually exclusive possibilities*, one and only one of which is true in a given circumstance or a given realization of an experiment.

- Example. A die (singular of dice) is rolled. The number of spots  $s$  is one of the six possibilities  $\{1, 2, 3, 4, 5, 6\}$ . In any given case one and only one of these appears.

★ The event algebra  $\mathcal{E}$  is a collection of certain subsets of elements of the sample space  $\mathcal{S}$  which form a *Boolean algebra*: closed under complements, intersections, unions. It always includes  $\mathcal{S}$  and the empty set  $\emptyset$ .

- The complement  $A^c$  of  $A \subset \mathcal{S}$  is all elements of  $\mathcal{S}$  not in  $A$ .

- Simplest situation to think about:  $\mathcal{S}$  is finite, or at most countably infinite, and  $\mathcal{E}$  consists of *all* subsets of  $\mathcal{S}$ , including  $\mathcal{S}$  itself and the empty set  $\emptyset$ . Sometimes it is convenient to choose a smaller collection of subsets.

- For the die: The event algebra contains  $2^6 = 64$  subsets of  $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$  if one includes  $\mathcal{S}$  and the empty set  $\emptyset$ . Another possibility:  $\mathcal{E}$  consists of the four subsets:  $\emptyset$ ,  $\{1, 3, 5\}$ ,  $\{2, 4, 6\}$ ,  $\mathcal{S}$ . That is, one is only concerned with whether the number of spots is even or odd.

- Exercise. Check that these four subsets of  $\mathcal{S}$  form a Boolean algebra.

★ Probabilities  $\mathcal{P}$ : One assigns real numbers between 0 and 1 to the elements  $\mathcal{E}$  following certain rules.

- Simplest situation, which will work for most of what we want to do. Let  $\mathcal{S}$  be finite or at most countably infinite;  $\mathcal{E}$  consists of all subsets of  $\mathcal{S}$ . Then for each  $s \in \mathcal{S}$  let  $p_s$  be a nonnegative

---

<sup>1</sup>M. Le Bellac, *Quantum Physics* (Cambridge, 2006). The quotations are from pp. 22 and 23.

real number, and choose these numbers such that

$$\sum_s p_s = 1. \tag{1}$$

- Then for any  $E \subset \mathcal{S}$  define its probability as

$$\Pr(E) = \sum_{s \in E} p_s \tag{2}$$

- Important! Probability theory *as such* does not assign a probability distribution. It merely specifies the rules when such a distribution is given.

- In the case of a die there is nothing that says  $p_s = 1/6$  for each  $s$  between 1 and 6; there are plenty of other possibilities. And some of these probably provide a better description of some of the dice that have shown up in Las Vegas. . . .

- Probabilities enter science as parameters in various models, and there is no “law of nature” that prescribes their values. There has been a lot of discussion of how to assign probabilities.

- Quantum theory is somewhat exceptional in that there are certain cases (to be discussed later) in which some probabilities enter the theory as axioms, and in this sense certain probabilities emerge from “laws of nature.”

## 2.2 Quantum

★ A *quantum sample space*  $\mathcal{S}$  is always a (projective) *decomposition of the identity operator*  $I$  for an appropriate quantum Hilbert space: a collection of projectors  $\{P_j\}$  which sum to  $I$ .

- The simplest situation is that of an orthonormal basis  $\{|b_j\rangle\}$  with  $P_j = [b_j] = |b_j\rangle\langle b_j|$ .

- It is convenient to employ the abbreviation  $[\psi]$  for the projector  $|\psi\rangle\langle\psi|$  onto the ray corresponding to the normalized ket  $|\psi\rangle$ .

- Example. Spin half.  $\{[z^+], [z^-]\}$

- The fact that  $P_j P_k = 0$  for  $j \neq k$  means that the properties corresponding to the different projectors in the decomposition are *mutually exclusive*; only one of them can be true in any given situation or realization of an experiment. The fact that the projectors sum to  $I$  means that one of them is true, since  $I$  is the property that is always true.

- Side remark. A “classical” object such as a die can in principle be described in quantum mechanical terms, and then the different possibilities when it has been tossed, different numbers of spots, correspond to orthogonal projectors on an appropriate Hilbert space. Their adding up to the identity is more subtle: they sum to a big projector whose physical significance is that all reasonable possibilities are included: the die is on top of the table, it has not melted, etc. In this sense classical probability theory follows the quantum rules, as one would expect if the universe is fully quantum mechanical.

- ★ Given  $\mathcal{S}$ , we shall employ an *event algebra*  $\mathcal{E}$  consisting of all projectors of the form

$$P = \sum_j \pi_j P_j, \tag{3}$$

where each  $\pi_j$  is either 0 or 1. Thus if there are  $n$  projectors in the decomposition  $\mathcal{S}$  there are  $2^n$  projectors in the event algebra  $\mathcal{E}$ .

□ Exercise. Show that if  $P$  is a projector in  $\mathcal{E}$ , then its complement  $I - P$  is in  $\mathcal{E}$ , and if  $P$  and  $Q$  are two projectors in  $\mathcal{E}$ , so is their product  $PQ = QP$ . With these operations  $\mathcal{E}$  is a Boolean algebra.

★ Two quantum sample spaces  $\mathcal{S} = \{P_j\}$  and  $\mathcal{T} = \{Q_k\}$  for the same system (the same Hilbert space) are *compatible* if  $P_j Q_k = Q_k P_j$  for every  $j$  and  $k$ ; otherwise they are *incompatible*.

◦ Example: The sample spaces  $\mathcal{S} = \{[z^+], [z^-]\}$  and  $\mathcal{T} = \{[x^+], [x^-]\}$  for a spin-half particle are incompatible.

• In the compatible case one can always form a third sample space  $\mathcal{R} = \{R_l\}$  (which might be identical to either  $\mathcal{S} = \{P_j\}$  or  $\mathcal{T} = \{Q_k\}$ ) by using the decomposition of the identity consisting of all products  $P_j Q_k$ , throwing away any duplicates and also all products that are 0. One calls  $\mathcal{R}$  the *coarsest common refinement* of  $\mathcal{S}$  and  $\mathcal{T}$ . The event algebra corresponding to  $\mathcal{R}$  contains all the projectors in both of the event algebras generated by  $\mathcal{S}$  and  $\mathcal{T}$ , along with (in general) others as well.

□ Exercise. Show that if  $\mathcal{S} \neq \mathcal{T}$  refer to spin half (a two-dimensional Hilbert space), they will typically be incompatible with each other.

□ Exercise. Construct a simple example of two nontrivial, compatible sample spaces  $\mathcal{S} \neq \mathcal{T}$  on a three-dimensional Hilbert space, and find the largest common refinement  $\mathcal{R}$ . (A trivial sample space is the trivial decomposition consisting of nothing but  $I$ .)

• The incompatible case arises in quantum, but not classical, mechanics, and here the single-framework rule kicks in: Do not try and combine results for incompatible quantum sample spaces. Trying to do so leads to paradoxes.

★ Once a sample space and the corresponding event algebra have been specified, as in Sec. 2, quantum probabilities behave in the same way as classical probabilities. The tricky issue is how to choose the sample space, and the issue of combining or not combining sample spaces. But if one has a single sample space and uses its corresponding event algebra there will be no problems.

• In particular, to each projector  $P_j$  in the decomposition one assigns a non-negative probability  $p_j$  in such a way that  $\sum_j p_j = 1$ . To a projector of the form (3) assign a probability

$$\Pr(P) = \sum \pi_j p_j. \quad (4)$$

• Just as in ordinary or classical probabilities, there is no general rule for assigning probabilities to a quantum sample space or the corresponding event algebra, though in some situations (which we will discuss later) the Born rule and its generalizations specify some aspects of how to do it.

### 3 Random and Physical Variables

#### 3.1 Classical

★ In ordinary “classical” probability theory a *random variable*  $V$  is a real-valued function  $V(s)$  defined on the sample space  $\mathcal{S}$ .

★ A particular case is that of the *indicator function* of the set  $E \subset \mathcal{S}$ , which can be (fairly safely) denoted by the same symbol  $E$ , the function defined by

$$E(s) = \begin{cases} 1 & \text{if } s \in E, \\ 0 & \text{if } s \notin E. \end{cases} \quad (5)$$

• If the set  $E$  consists of the single point  $s = r$ , i.e.,  $E = \{r\}$ , we shall use the special notation  $J_r(s)$  for the corresponding indicator function:

$$J_r(s) = \delta_{rs} = \begin{cases} 1 & \text{if } s = r, \\ 0 & \text{if } s \neq r. \end{cases} \quad (6)$$

- Assuming that  $\mathcal{S}$  is finite or at most countably infinite, any random variable can be written as a sum of indicators multiplied by function values, in two slightly different ways:

$$V(s) = \sum_r v_r J_r(s) = \sum_j v'_j P_j(s). \quad (7)$$

Here  $v_r = V(r)$  is just the value which  $V(s)$  takes when  $s = r$ . The second sum is over all the *distinct* values  $v'_j$  that  $V(s)$  can take, with  $j$  some appropriate index or label, and

$$P_j(s) = \begin{cases} 1 & \text{if } V(s) = v'_j, \\ 0 & \text{if } V(s) \neq v'_j. \end{cases} \quad (8)$$

is the indicator of the set of all  $s$  where  $V(s)$  takes on the particular value  $v'_j$ .

□ Exercise. Work this out in the case in which  $S = \{1, 2, 3\}$  and  $V(1) = 1$ ,  $V(2) = 2$ , and  $V(3) = 2$ .

★ The probability that  $V(s)$  takes on the value  $v$  is given by

$$\Pr(V = v) = \Pr(P_j) = \langle P_j \rangle \quad (9)$$

when  $v'_j = v$ ;  $\langle \rangle$  is defined next.

★ A very important concept is that of the *average* of a random variable  $V$ , also called its *expectation* or *expectation value*:

$$\langle V \rangle = \sum_s p_s V(s). \quad (10)$$

◦ If  $\mathcal{S}$  is infinite the average may not be defined; we restrict our considerations to cases in which it is defined (i.e., the sum is absolutely convergent).

- The average value of an indicator function  $E(s)$  is equal to the probability of the corresponding set or event  $E$ :

$$\langle E \rangle = \Pr(E) = \sum_{s \in E} p_s. \quad (11)$$

□ Exercise. Verify this

★ The *variance* of  $V$  is

$$\text{Var}(V) = (\Delta V)^2 = \langle V^2 \rangle - (\langle V \rangle)^2 = \langle V^2 - \langle V \rangle^2 \rangle, \quad (12)$$

and the positive square root  $\Delta V$  of the variance is known as the *standard variation*. This is useful for indicating the “width” of the probability distribution. Note that  $\Delta V$  has the same dimensions as  $V$  when the latter is a dimensioned quantity (e.g., energy).

★ In classical mechanics, physical variables such as energy, and the different components of position, momentum, and angular momentum are examples of real valued functions on the phase space, and we shall call them *physical variables*, in contrast to *physical properties*, a term which is used here for subsets of the phase space. If, as in (classical) statistical mechanics, the phase space is regarded as a sample space, then physical variables are examples of random variables of the form used in probability theory.

- The phase space consists of an uncountably infinite number of points, and this leads to various technical complications. One way of getting around them is to imagine that the phase space has been coarse grained by dividing it up into a countable collection of cells of equal size, or even a finite collection of cells if one lets some of the cells be of infinite size. This sometimes provides a more helpful analogy with quantum theory than does a continuous phase space.

### 3.2 Quantum

★ In quantum theory the counterpart of a random variable or a classical physical variable is a Hermitian operator  $V$  that assigns a real number to every element of a decomposition of the identity  $\mathcal{S} = \{P_j\}$  in the sense that

$$V = \sum_j v'_j P_j, \quad (13)$$

where we assume that the eigenvalues on the right side are distinct:  $v'_j \neq v'_k$  for  $j \neq k$ .

◦ One could, of course, use some refinement of  $\{P_j\}$  and rewrite  $V$  in terms of a sum over these projectors, in which case the same eigenvalue would (in general) appear more than once. For example,

$$V = \sum_k v_k [b_k] = \sum_k v_k |b_k\rangle\langle b_k|, \quad (14)$$

where  $\{|b_k\rangle\}$  is an orthonormal basis of the Hilbert space composed of eigenkets of  $V$ .

• Since the  $v_j$  are, by assumption, real,  $V = V^\dagger$  is a Hermitian operator. Similarly, any Hermitian operator  $V$ , which is to say the operator for *any* quantum *physical variable*, can be written in the form (13) or (14), where the coefficients on the right side of the equation are its eigenvalues.

– In quantum theory a physical variable is often referred to as an *observable*. This is a convenient technical term if one ignores its historical origin in an outdated theory of measurements.

★ There is a unique decomposition of the identity associated with a physical variable  $V$  such that the expansion (13) contains only one term for each distinct eigenvalue. We shall say that this is the decomposition or sample space *corresponding* to or *generated by* the operator in question.

◦ If some of the projectors are of rank (trace) greater than 1, then there will also exist an alternative way of writing  $V$  using a finer decomposition of the identity, as in (14).

• In particular,  $P_j$  in (13), or the subspace onto which it projects, corresponds to the property that  $V$  takes on the value  $v'_j$ , in much the same way as the subset of the classical sample space for which  $P_j(s)$  in (8) is the indicator function consists of and contains all the points where the random variable  $V(s)$  takes on the value  $v'_j$  in (7).

★ One says that a quantum physical variable or observable  $V$  *takes on* or *has* the value  $v$  in the quantum state  $|\psi\rangle$  provided  $|\psi\rangle$  is an eigenstate of  $V$  with eigenvalue  $v$ , i.e.,

$$V|\psi\rangle = v|\psi\rangle, \quad (15)$$

which is so if and only if  $|\psi\rangle$  lies in the corresponding subspace  $P_j$ , (13), for which  $v'_j = v$

• The probability that  $V$  takes on the value  $v$  is the probability assigned to the corresponding subspace of the Hilbert space, or its projector:

$$\Pr(V = v'_j) = \Pr(P_j) = p_j. \quad (16)$$

– While the notation  $V = v'_j$ , equating an operator to a scalar, may seem a bit odd, its significance should be clear in view of the preceding discussion.

★ The average value of  $V$  is then

$$\langle V \rangle = \sum_j v'_j p_j. \quad (17)$$

• This same average can also be written in the alternative form

$$\langle V \rangle = \text{Tr}(\rho V), \quad (18)$$

where

$$\rho = \sum_j p_j P_j / \text{Tr}(P_j) \quad (19)$$

is a positive operator of trace equal to 1, thus a density operator.

◦ Note that  $\rho$  and  $V$  commute, since they are written as sums of projectors from the same decomposition of the identity. We will later encounter cases in which  $\rho$  and  $V$  do not commute, and this will require a separate discussion.

★ The *variance*  $\Delta V^2$  and the *standard deviation*  $\Delta V$  are defined in the quantum case using (12), just as in the case of a classical random variable.

## 4 Several Random (Physical) Variables

### 4.1 Classical

• One often makes use of several random variables, which is to say several real-valued functions on the same sample space  $\mathcal{S}$ , again assumed to be finite or at most countably infinite. The essential ideas can be illustrated by by considering two random variables  $V(s)$  and  $W(s)$ .

★ The joint probability distribution for  $V$  and  $W$ ,  $\Pr(V = v, W = w)$  or  $\Pr(v, w)$  for short, is defined to be

$$\Pr(v, w) = \langle EF \rangle = \sum_{s \in E \cap F} p_s, \quad (20)$$

where  $E$  is the subset of  $\mathcal{S}$  where  $V(s) = v$ , or its indicator, and  $F$  the set (indicator) where  $W(s) = w$ , and the product  $EF$  is the random variable  $E(s)F(s)$ , the indicator for the intersection  $E \cap F$ , where it should cause no confusion if we use the same symbol for a subset of  $\mathcal{S}$  and for its indicator function. Read  $(v, w)$  or  $(V = v, W = w)$  as “ $V = v$  AND  $W = w$ .”

• Note how the “marginal” distribution  $\Pr(V = v)$  defined earlier in (9), and its counterpart for  $W$ , can be obtained from the joint distribution:

$$\Pr(V = v) = \sum_w \Pr(v, w); \quad \Pr(W = w) = \sum_v \Pr(v, w) \quad (21)$$

◦ Since we are dealing with a sample space that is at most countable, the possible values of  $V$  and  $W$  are discrete, and the sums in (21) have an obvious meaning. When the sample space is continuous (outside the scope of these notes) one may have to replace sums with integrals.

★ Conditional probabilities.

• Let  $A$  and  $B$  denote any two subsets of the sample space  $\mathcal{S}$ . Provided  $\Pr(B) > 0$ , define the *conditional probability of  $A$  given  $B$*  by

$$\Pr(A | B) = \Pr(A \cap B) / \Pr(B) = \Pr(A, B) / \Pr(B), \quad (22)$$

where either  $\Pr(A, B)$  or  $\Pr(A \cap B)$  is the probability of  $A$  AND  $B$ , the probability of the intersection of the corresponding subsets of  $\mathcal{S}$ . This can also be written as  $\Pr(AB)$  if one thinks of  $A$  and  $B$  as the corresponding indicators.

– If  $\Pr(B) = 0$  the conditional  $\Pr(A | B)$  is undefined. This is a technical annoyance which the reader should be aware of (and we will not emphasize further). In some sense is of no practical significance, since an event  $B$  with zero probability occurs with zero probability (i.e., never).

• One can also write (22) in the equivalent form

$$\Pr(A, B) = \Pr(A | B) \Pr(B). \quad (23)$$

- It is worthwhile memorizing both (22) and (23).
- Conditional probabilities for two random variables  $V$  and  $W$ :

$$\Pr(v|w) = \Pr(V = v | W = w) = \Pr(v, w) / \Pr(w), \text{ or } \Pr(v, w) = \Pr(v|w) \Pr(w). \quad (24)$$

★ Conditional probabilities play a fundamental role in probabilistic reasoning. So it is important to learn to use them properly and effectively.

- As a function of its *first* argument, with second argument fixed, a conditional probability behaves like any other probability.
  - For example,  $\sum_v \Pr(v|w) = 1$ .
  - Exercise. Show that if we define

$$\bar{p}_s := \Pr(s | B) \quad (25)$$

then

$$\Pr(A | B) = \sum_{s \in A} \bar{p}_s \quad (26)$$

- Case of a die. Let us assume it is honest, it has just been rolled, and you have not seen it. So you don't know the value of  $s$ , and you assume a probability  $1/6$  for each outcome.
  - Now you are told that the number of spots is even. Given that information, you revise your probability distribution: you now assign a probability of 0 to  $s = 1, 3,$  and  $5$ ; and  $1/3$  to  $s = 2, 4,$  and  $6$ .

- Exercise. Show that these new probabilities are those in (25) if  $B$  is the subset “ $s$  is even.”
- Think of the condition probability  $\Pr(A | B)$  as obtained by setting to zero the probabilities for all  $s$  that are outside  $B$ . Then multiply each  $p_s$  for  $s \in B$  by the same amount,  $1/\Pr(B)$ , to obtain a new probability distribution

★ Two events  $A$  and  $B$  are said to be *statistically independent* provided

$$\Pr(A, B) = \Pr(A) \Pr(B) \text{ or } \Pr(A | B) = \Pr(A) \text{ or } \Pr(B | A) = \Pr(B). \quad (27)$$

□ Exercise. Show that each equality in (27) implies the other two (assuming that the conditional probabilities exist, i.e.,  $\Pr(A) > 0, \Pr(AB) > 0$ ).

- While each equality in (27) implies the other, it is worthwhile memorizing both the first and the second, as the intuitive notions are a bit different.
- Two random variables  $V$  and  $W$  are said to be statistically independent provided

$$\Pr(v, w) = \Pr(v) \Pr(w) \quad (28)$$

for every possible  $v$  and  $w$ .

□ Exercise. Suppose both  $V$  and  $W$  take on only two values, 0 or 1, i.e., they are indicators. It then looks as if checking (28) means checking four independent conditions. Show that checking one of them suffices; the other three will then follow. Can you relate this to (27)?

• One can think of independent random variables in the following way. Additional information will typically allow you to update a probability distribution to something that is more precise or informative. (See above example of rolling a die where you are told that the number of spots is even.) However, there are cases in which additional information tells you nothing at all about some event or random variable you happen to be interested in.



◦ Example using an honest die: let  $A$  be “ $s > 3$ ” and  $B$  be “ $s = 1$  or  $6$ ”.

□ Exercise. Work it out assuming  $p_s = 1/6$  independent of  $s$ . But what about some other probability distribution? Give an example with  $p_s \neq 1/6$  for which  $A$  and  $B$  are statistically independent, and another example where they are not statistically independent.

★ The extension of these ideas to three or more events or random variables is fairly straightforward, so we won’t give details.

• For example, let  $U$ ,  $V$ , and  $W$  be three random variables. Statistical independence is then the requirement

$$\Pr(u, v, w) = \Pr(u) \Pr(v) \Pr(w) \quad (29)$$

in an obvious notation.

◦ Also,

$$\Pr(v, w) = \sum_u \Pr(u, v, w) \quad (30)$$

$$\Pr(v, w | u) = \Pr(u, v, w) / \Pr(u), \quad (31)$$

assuming  $\Pr(u) > 0$ .

## 4.2 Quantum

★ In quantum theory the situation with two or more physical variables or observables is precisely the same, or at least closely analogous, to that in ordinary probability theory *as long as the operators commute*. When they do not commute the single framework rules forbids combining them.

• Simultaneous diagonalization. If  $\{V, W, X, \dots\}$  are a collection of *commuting* normal operators,  $VW = WV$  and so forth for all pairs, then they can be simultaneously diagonalized using the same orthonormal basis. That is, there is an orthonormal basis such that each basis ket is an eigenket of each of the operators.

• This basis corresponds to a decomposition of the identity using projectors of rank 1. Sometimes it is possible to use a coarser decomposition if one has some collection of basis kets such that each operator in the collection assigns the same eigenvalue to each of these kets. Then one can replace the individual dyads with a projector onto the space they span.

◦ There is always a coarsest decomposition of the identity  $\{P_j\}$  such that all the commuting operators can be written in terms of projectors from this set.

• If somehow probabilities  $p_j$  have been assigned to each  $P_j$ , with (of course) sum equal to 1, the situation is just like the classical case: there is a joint probability distribution, which then leads to conditional probabilities of the usual sort, which can be interpreted in the usual way.

◦ Statistical independence is defined in the obvious way.

★ If, on the other hand,  $V$  and  $W$  are two Hermitian operators that do *not* commute, they are *incompatible* observables (physical variables). and discussions of their probabilities have to be carried out in separate frameworks, and cannot be combined.

• However, the situation illustrated by the following example sometimes arises:

$$V = v_1 P_1 + v_2 P_2 + v_3 P_3, \quad W = w_1 P_1 + w_2 Q_2 + w_3 Q_3, \quad (32)$$

where  $\{P_1, P_2, P_3\}$  and  $\{P_1, Q_2, Q_3\}$  are incompatible decompositions of the identity because while  $P_2 + P_3 = Q_2 + Q_3$ ,  $P_2 Q_2 \neq Q_2 P_2$ , etc.

◦ If one then adopts a decomposition  $\{P_1, I - P_1\}$  of the identity with just two projectors this can form a sample space which, while not as fine grained as one would like, allows one speak sensibly

and simultaneously (i.e., using a single framework) about the events  $V = v_1$ ,  $V \neq v_1$ ,  $W = w_1$ , and  $W \neq w_1$ , where by “ $V \neq v_1$ ” one means what is represented by the projector  $I - P_1$ , and not that “it is either the case that  $V = v_2$  or else it is the case that  $V = v_3$ ”, which would then require that  $P_2$  and  $P_3$  both belong to the framework.

## 5 Born Rule

### 5.1 Pre-probabilities

★ There is a uniquely quantum mechanical way of assigning probabilities, with no classical analog, known as the *Born rule*. We introduce it here simply as a calculational device. Later on we will see it should be applied in various physical situations, in particular involving time development. The connection between the Born rule and a certain class of idealized measurements is discussed in Sec. 5.2.

★ Let  $|\psi\rangle$  be a normalized ket,  $\langle\psi|\psi\rangle = 1$ . We shall call it a *pre-probability* when it is used to generate a set of probabilities  $p_j$  for the elements  $P_j$  of a quantum sample space or decomposition of the identity  $\{P_j\}$ , see Sec. 2.2, using the formula

$$p_j = \langle\psi|P_j|\psi\rangle = \text{Tr}([\psi]P_j). \quad (33)$$

□ Exercise. Why are these  $p_j$  nonnegative, and why do they sum to 1?

• In particular, if the sample space is associated with an orthonormal basis  $\{|b_j\rangle\}$ , the Born rule assigns probability a

$$p_j = |\langle b_j|\psi\rangle|^2 = |\langle\psi|b_j\rangle|^2 = \text{Tr}([\psi][b_j]) \quad (34)$$

to the property  $[b_j]$  corresponding to  $|b_j\rangle$  (i.e., to the ray which contains the ket  $|b_j\rangle$ ).

◦ If  $|\psi\rangle$  is not normalized it can still be used as a pre-probability; simply replace it with  $|\bar{\psi}\rangle = |\psi\rangle/\sqrt{\langle\psi|\psi\rangle}$  in (33) or (34).

★ A density operator  $\rho$ , a positive operator with trace equal to 1, can also serve as a pre-probability, in which case (33) is to be replaced with

$$p_j = \text{Tr}(\rho P_j). \quad (35)$$

• Note that the projector  $[\psi]$  is a positive operator with trace 1, and hence a special kind of density operator. In that case (35) is obviously identical to the right side of (33).

★ It is important to distinguish a *pre-probability* from a quantum *property*. In the simplest situation both are represented by kets, so it is easy to confuse them. The fundamental difference is that a property is something real, whereas a pre-probability, which is used to generate probabilities, is no more real than a probability as used in classical physics. It is a mathematical abstraction.

◦ To be sure, a vector in the Hilbert space is itself a mathematical abstraction, not a physical reality. It can, however, *represent* a physical property in much the same way that three numbers in an appropriate coordinate system can represent the present position of the center of mass of the planet Jupiter. A probability cannot correspond to a physical property in the same way (unless the probability is 0 or 1, in which case it becomes a certainty).

★ Example. Let  $|\psi\rangle = |x^+\rangle$  be the pre-probability, and  $\{|z^+\rangle, |z^-\rangle\}$  the sample space. Then since  $|\langle x^+|z^+\rangle| = |\langle x^+|z^-\rangle| = 1/\sqrt{2}$ , we conclude that  $\text{Pr}(z^+) = \text{Pr}(z^-) = 1/2$ .

• If, on the other hand, we use the same pre-probability  $|\psi\rangle = |x^+\rangle$  but a different sample space  $\{|x^+\rangle, |x^-\rangle\}$ , then  $\text{Pr}(x^+) = 1$ ,  $\text{Pr}(x^-) = 0$ .

!! Beware! Since the same pre-probability is used in the two instances just discussed, one may be tempted to combine the results and say that: “the probability to  $x^+$  is 1, the probabilities of  $z^+$  and  $z^-$  are 1/2, and the probability of  $x^-$  is 0.” The trouble with this statement is that it suggests that  $[x^+]$ ,  $[z^+]$ ,  $[z^-]$ , and  $[x^-]$  are all elements of a single event algebra, and one is comparing them using a common probability distribution. But  $[x^+]$  and  $[z^+]$  are incompatible (in the quantum sense), and it makes no sense to compare their probabilities.

## 5.2 Measurements

★ The Born rule is often stated in quantum textbooks in the following way, where to make it definite we consider the case of a spin half particle. Suppose that the normalized quantum state is  $|\psi\rangle$ . The textbook might say:

M. “If  $S_z$  is measured the probability of finding  $S_z = 1/2$  (in units of  $\hbar$ ) is  $|\langle z^+|\psi\rangle|^2$ , and the probability of finding  $S_z = -1/2$  is  $|\langle z^-|\psi\rangle|^2$ .”

- This can be interpreted in the following way. Mentioning a particular quantum observable or physical variable, in this case  $S_z$ , implicitly defines a decomposition of the identity in terms of projectors onto its eigenspaces, as in (13). In the case at hand that decomposition is  $\{[z^+], [z^-]\}$  corresponding to the orthonormal basis  $\{|z^+\rangle, |z^-\rangle\}$ . Next comes the notion of an *ideal measurement* in which the measurement outcome is indicated on the measuring instrument by clearly distinguished macroscopic (“classical”) states of affairs. In a tradition that goes back much earlier than modern electronics one speaks of the different outcomes as distinct positions of a *pointer*. An ideal measurement is one in which the apparatus pointer position accurately reflects the state of affairs that existed just before the measurement was carried out. The probabilities referred to in M are, strictly speaking, the probabilities that the measurement pointer will be in one position or the other. However, in an ideal measurement the probabilities for the pointer positions are the same as the probabilities of the properties the instrument was designed to measure. So we end up in the end with probabilities of certain quantum properties.

★ There are two reasons for this strange circumlocution with its reference to “measurements”.

- The first is tradition: the textbook writer learned quantum mechanics from a previous book written by someone who learned quantum mechanics from a previous book written by ... by John von Neumann with the title *Mathematische Grundlagen der Quantenmechanik*. Von Neumann was a bit confused about this aspect of quantum theory, and his confusion has been passed down to his intellectual grandchildren.

- Second, measurements are being used to cover up conceptual difficulties of the quantum world. Hilbert space quantum mechanics allows one to speak in a sensible way about  $S_x$  for a spin half particle, or about  $S_z$ , or any other component, but combining descriptions of angular momentum components referring to different directions in space is not possible. Rather than facing up to this difficulty directly, what textbooks state is that there is impossible to simultaneously *measure* both  $S_x$  and  $S_z$  for a spin half particle.

- This statement, that  $S_x$  and  $S_z$  cannot be simultaneously measured, is correct. What textbooks omit is the explanation: what does not exist cannot be measured! There just is nothing in the quantum Hilbert space of a spin half particle that could represent the particle having both a value (necessarily  $\pm 1/2$ ) of  $S_x$  and a value of  $S_z$ . Even very competent experimentalists cannot measure what is not there! (Indeed, this is one thing that distinguishes them from less competent colleagues.)

- Later on we will discuss the measurement process, using a simplified but fully consistent quantum mechanical description that includes the apparatus as well as the measured system.

★ Until then, the reader should interpret textbook references to “measurement” as a way of

identifying a quantum sample space:

◦ Textbook statement: “Given that the quantum state is  $|\psi\rangle$ , what is the probability that the measurement of  $S_z$  will yield the value of  $+1/2$ ?”

◦ Translation: “Given the quantum state  $|\psi\rangle$  understood as a pre-probability, and that  $S_z = ([z^+] - [z^-])/2$  so the sample space consists of  $[z^+]$  and  $[z^-]$  (the projectors onto the eigenstates  $|z^+\rangle$  and  $|z^-\rangle$  of  $S_z$ ), what is the probability of  $|z^+\rangle$ ?”