Chapter 5

Probabilities and Physical Variables

5.1 Classical Sample Space and Event Algebra

Probability theory is based upon the concept of a *sample space* of mutually exclusive possibilities, one and only one of which actually occurs, or is true, in any given situation. The elements of the sample space are sometimes called *points* or *elements* or *events*. In classical and quantum mechanics the sample space usually consists of various possible states or properties of some physical system. For example, if a coin is tossed, there are two possible outcomes: H (heads) or T (tails), and the sample space $\mathcal{S}$ is {H,T}. If a die is rolled, the sample space $\mathcal{S}$ consists of six possible outcomes: $s = 1, 2, 3, 4, 5, 6$. If two individuals A and B share an office, the occupancy sample space consists of four possibilities: an empty office, A present, B present, or both A and B present.

Associated with a sample space $\mathcal{S}$ is an *event algebra* $\mathcal{B}$ consisting of subsets of elements of the sample space. In the case of a die, "s is even" is an event in the event algebra. So are "s is odd", "s is less than 4", and "s is equal to 2." It is sometimes useful to distinguish events which are elements of the sample space, such as $s = 2$ in the previous example, and those which correspond to more than one element of the sample space, such as "s is even". We shall refer to the former as *elementary* events and to the latter as *compound* events. If the sample space $\mathcal{S}$ is finite and contains n points, the event algebra contains 2^n possibilities, including the entire sample space $\mathcal{S}$ considered as a single compound event, and the empty set $\emptyset$. For various technical reasons it is convenient to include $\emptyset$, even though it never actually occurs: it is the event which is always false. Similarly, the compound event $\mathcal{S}$, the set of all elements in the sample space, is always true. The subsets of $\mathcal{S}$ form a *Boolean algebra* or *Boolean lattice* $\mathcal{B}$ under the usual set-theoretic relationships: The *complement* $\sim\mathcal{E}$ of a subset $\mathcal{E}$ of $\mathcal{S}$ is the set of elements of $\mathcal{S}$ which are not in $\mathcal{E}$. The *intersection* $\mathcal{E} \cap \mathcal{F}$ of two subsets is the collection of elements they have in common, whereas their *union* $\mathcal{E} \cup \mathcal{F}$ is the collection of elements belonging to one or the other, or possibly both.

The phase space of a classical mechanical system is a sample space, since one and only one point in this space represents the actual state of the system at a particular time. Since this space contains an uncountably infinite number of points, one usually defines the event algebra not as the collection of all subsets of points in the phase space, but as some more manageable collection, such as the Borel sets.

A useful analogy with quantum theory is provided by a *coarse graining* of the classical phase space, a finite or countably infinite collection of non-overlapping regions or *cells* which together cover the phase space. These cells, which in the notation of Ch. 4 represent properties of the physical system, constitute a sample space $\mathcal{S}$ of mutually exclusive possibilities, since a point γ in the phase space representing the state of the system at a particular time will be in one and only one cell, making this cell a true property, whereas the properties corresponding to all of the other cells in the sample are false. (Note that individual points in the phase space are not, in and of themselves, members of $\mathcal{S}$.) The event algebra $\mathcal{B}$ associated with this coarse graining consists of collections of one or more cells from the sample space, along with the empty set and the collection of all the cells. Each event in $\mathcal{B}$ is associated with a physical property corresponding to the set of all points in the phase space lying in one of the cells making up the (in general compound) event. The negation of an event $\mathcal{E}$ is the collection of cells which are in $\mathcal{S}$ but not in $\mathcal{E}$, the conjunction of two events $\mathcal{E}$ and $\mathcal{F}$ is the collection of cells which they have in common, and their disjunction the collection of cells belonging to $\mathcal{E}$ or to $\mathcal{F}$ or to both.

As an example, consider a one-dimensional harmonic oscillator whose phase space is the x, p plane. One possible coarse graining consists of the four cells

$$x \geq 0,\, p \geq 0; \quad x < 0,\, p \geq 0; \quad x \geq 0,\, p < 0; \quad x < 0,\, p < 0; \tag{5.1}$$

that is, the four quadrants defined so as not to overlap. Another coarse graining is the collection $\{C_n\}, n = 1, 2, \ldots$ of cells

$$C_n :\ (n-1)E_0 \leq E < nE_0 \tag{5.2}$$

defined in terms of the energy E, where $E_0 > 0$ is some constant. Still another coarse graining consists of the rectangles

$$D_{mn} :\ mx_0 < x \leq (m+1)x_0,\ np_0 < p \leq (n+1)p_0, \tag{5.3}$$

where $x_0 > 0$, $p_0 > 0$ are constants, and m and n are any integers.

As in Sec. 4.1, we define the *indicator* or *indicator function* E for an event $\mathcal{E}$ to be the function on the sample space which takes the value 1 space which is in the set $\mathcal{E}$, and 0 (false) on all other elements:

$$E(s) = \begin{cases} 1 & \text{for } s \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.4}$$

The indicators form an algebra under the operations of negation ($\sim E$), conjunction ($E \wedge F$), and disjunction ($E \vee F$), as discussed in Secs. 4.4 and 4.5:

$$\begin{aligned} \sim E &= \tilde{E} = I - E, \\ E \wedge F &= EF, \\ E \vee F &= E + F - EF, \end{aligned} \tag{5.5}$$

where the arguments of the indicators have been omitted; one could also write $(E \wedge F)(s) = E(s)F(s)$, etc. Obviously $E \wedge F$ and $E \vee F$ are the counterparts of $\mathcal{E} \cap \mathcal{F}$ and $\mathcal{E} \cup \mathcal{F}$ for the corresponding subsets of $\mathcal{S}$. We shall use the terms "event algebra" and "Boolean algebra" for either the algebra of sets or the corresponding algebra of indicators.

Associated with each element r of a sample space is a special indicator P_r which is zero except at the point r:

$$P_r(s) = \begin{cases} 1 & \text{if } s = r, \\ 0 & \text{if } s \neq r. \end{cases} \tag{5.6}$$

Indicators of this type will be called *elementary* or *minimal*, and it is easy to see that

$$P_r P_s = \delta_{rs} P_s. \tag{5.7}$$

The vanishing of the product of two elementary indicators associated with distinct elements of the sample space reflects the fact that these events are mutually exclusive possibilities: if one of them occurs (is true), the other cannot occur (is false), since the zero indicator denotes the "event" which never occurs (is always false). An indicator R on the sample space corresponding to the (in general compound) event $\mathcal{R}$ can be written as a sum of elementary indicators,

$$R = \sum_{s \in \mathcal{S}} \pi_s P_s, \tag{5.8}$$

where π_s is equal to 1 if s is in $\mathcal{R}$, and 0 otherwise. The indicator I, which takes the value 1 everywhere, can be written as

$$I = \sum_{s \in \mathcal{S}} P_s, \tag{5.9}$$

which is (5.8) with $\pi_s = 1$ for every s.

5.2 Quantum Sample Space and Event Algebra

In Sec. 3.5 a decomposition of the identity was defined to be an orthogonal collection of projectors $\{P_j\}$,

$$P_j P_k = \delta_{jk} P_j, \tag{5.10}$$

which sum to the identity

$$I = \sum_j P_j. \tag{5.11}$$

Any decomposition of the identity of a quantum Hilbert space $\mathcal{H}$ can be thought of as a *quantum sample space* of mutually exclusive properties associated with the projectors or with the corresponding subspaces. That the properties are mutually exclusive follows from (5.10), see the discussion in Sec. 4.5, which is the quantum counterpart of (5.7). The fact that the projectors sum to I is the counterpart of (5.9), and expresses the fact that one of these properties must be true. Thus the usual requirement that a sample space consist of a collection of mutually exclusive possibilities, one and only one of which is correct, is satisfied by a quantum decomposition of the identity.

The *quantum event algebra* $\mathcal{B}$ corresponding to the sample space (5.11) consists of all projectors of the form

$$R = \sum_j \pi_j P_j, \tag{5.12}$$

where each π_j is either 0 or 1; note the analogy with (5.8). Setting all the π_j equal to 0 yields the zero operator 0 corresponding to the property that is always false; setting them all equal to 1 yields the identity I, which is always true. If there are n elements in the sample space, there are 2^n elements in the event algebra, just as in ordinary probability theory. The *elementary* or *minimal* elements of $\mathcal{B}$ are the projectors $\{P_j\}$ which belong to the sample space, whereas the *compound* elements are those for which two or more of the π_j in (5.12) are equal to 1.

Since the different projectors which make up the sample space commute with each other, (5.10), so do all projectors of the form (5.12). And because of (5.10), the projectors which make up the event algebra $\mathcal{B}$ form a *Boolean algebra* or *Boolean lattice* under the operations of $\cap$ and $\cup$ interpreted as $\wedge$ and $\vee$; see (5.5), which applies equally to classical indicators and quantum projectors. Any collection of commuting projectors forms a Boolean algebra provided the negation $\tilde{P}$ of any projector P in the collection is also in the collection, and the product PQ $(= QP)$ of two elements in the collection is also in the collection. (Because of (4.50), these rules ensure that $P \vee Q$ is also a member of the collection, so this does not have to be stated as a separate rule.) Note that a Boolean algebra of projectors is a much simpler object (in algebraic terms) than the non-commutative algebra of all operators on the Hilbert space.

A *trivial* decomposition of the identity contains just one projector, I; *nontrivial* decompositions contain two or more projectors. For a spin-half particle, the only non-trivial decompositions of the identity are of the form

$$I = [w^+] + [w^-], \tag{5.13}$$

where w is some direction in space, such as the x axis or the z axis. Thus the sample space consists of two elements, one corresponding to $S_w = +1/2$ and one to $S_w = -1/2$. These are mutually exclusive possibilities: one and only one can be a correct description of this component of spin angular momentum. The event algebra $\mathcal{B}$ consists of the four elements: 0, I, $[w^+]$ and $[w^-]$.

Next consider a toy model, Sec. 2.5, in which a particle can be located at one of $M = 3$ sites, $m = -1, 0, 1$. The three kets $|-1\rangle, |0\rangle, |1\rangle$ form an orthonormal basis of the Hilbert space. A decomposition of the identity appropriate for discussing the particle's position contains the three projectors

$$[-1],\ \ [0],\ \ [1] \tag{5.14}$$

corresponding to the property that the particle is at $m = -1$, $m = 0$, and $m = 1$, respectively. The Boolean event algebra has $2^3 = 8$ elements: 0, I, the three projectors in (5.14), and three projectors

$$[-1] + [0],\ \ [0] + [1],\ \ [-1] + [1] \tag{5.15}$$

corresponding to compound events. An alternative decomposition of the identity for the same Hilbert space consists of the two projectors

$$[-1],\ \ [0] + [1], \tag{5.16}$$

which generate an event algebra with only $2^2 = 4$ elements: the projectors in (5.16) along with 0 and I.

Although the same projector $[0] + [1]$ occurs both in (5.15) and in (5.16), its physical interpretation or meaning in the two cases is actually somewhat different, and discussing the difference will throw light upon the issue raised at the end of Sec. 4.5 about the meaning of a quantum disjunction

$P \vee Q$. In (5.15), $[0] + [1]$ represents a *compound* event whose physical interpretation is that the particle is at $m = 0$ or at $m = 1$, in much the same way that the compound event $\{3, 4\}$ in the case of a die would be interpreted to mean that either $s = 3$ or $s = 4$ spots turned up. On the other hand, in (5.16) the projector $[0] + [1]$ represents an *elementary* event which cannot be thought of as the disjunction of two different possibilities. In quantum mechanics, each Boolean event algebra constitutes what is in effect a "language" out of which one can construct a quantum description of some physical system, and a fundamental rule of quantum theory is that a description (which may, but need not be couched in terms of probabilities) referring to a single system at a single time must be constructed using a single Boolean algebra, a single "language". (This is a particular case of a more general "single-framework rule" which will be introduced later on, and discussed in some detail in Ch. 16.) The language based on (5.14) contains among its elementary constituents the projector [0] and the projector [1], and its grammatical rules allow one to combine such elements with "and" and "or" in a meaningful way. Hence in this language "[0] or [1]" makes sense, and it is convenient to represent it using the projector $[0] + [1]$ in (5.15). On the other hand, the language based on (5.16) contains neither [0] nor [1]—they are not in the sample space, nor are they among the four elements which constitute its Boolean algebra. Consequently, in this somewhat impoverished language it is impossible to express the idea "[0] or [1]", because both [0] and [1] are meaningless constructs.

The reader may be tempted to dismiss all of this as needless nitpicking better suited to mathematicians and philosophers than to physical scientists. Is it not obvious that one can always replace the impoverished language based upon (5.16) with the richer language based upon (5.14), and avoid all this quibbling? The answer is that one can, indeed, replace (5.16) with (5.14) in appropriate circumstances; the process of doing so is known as "refinement", and will be discussed in Sec. 5.3 below. However, in quantum theory there can be many different refinements. In particular, a second and rather different refinement of (5.16) will be found below in (5.19). Because of the multiple possibilities for refinement, one must pay attention to what one is doing, and it is especially important to be explicit about the sample space ("language") that one is using. Shortcuts in reasoning which never cause difficulty in classical physics can lead to enormous headaches in quantum theory, and avoiding these requires that one take into account the rules which govern meaningful quantum descriptions.

As an example of a sample space associated with a continuous quantum system, consider the decomposition of the identity

$$I = \sum_n [\phi_n] \tag{5.17}$$

corresponding to the energy eigenstates of a quantum harmonic oscillator, in the notation of Sec. 4.3. The elementary event $[\phi_n]$ can be interpreted as the energy having the value $n + 1/2$ in units of $\hbar\omega$. These events are mutually exclusive possibilities: if the energy is 3.5, it cannot be 0.5 or 2.5, etc. The projector $[\phi_2] + [\phi_3]$ in the Boolean algebra generated by (5.17) means that the energy is equal to 2.5 or 3.5. If, on the other hand, one were to replace (5.17) with an alternative decomposition of the identity consisting of the projectors $\{([\phi_{2m}] + [\phi_{2m+1}]), m = 0, 1, 2 \ldots\}$, each projecting onto a two dimensional subspace of $\mathcal{H}$, $[\phi_2] + [\phi_3]$ could not be interpreted as an energy equal to 2.5 or 3.5, since states without a well-defined energy are also present in the corresponding subspace. See the preceding discussion of the toy model.

5.3 Refinement, Coarsening and Compatibility

Suppose there are two decompositions of the identity $\mathcal{E} = \{E_j\}$ and $\mathcal{F} = \{F_k\}$ with the property that each F_k can be written as a sum of one of more of the E_j. In such a case we will say that the decomposition $\mathcal{E}$ is a *refinement* of $\mathcal{F}$, or $\mathcal{E}$ is *finer* than $\mathcal{F}$, or $\mathcal{E}$ is obtained by *refining* $\mathcal{F}$. Equivalently, $\mathcal{F}$ is a *coarsening* of $\mathcal{E}$, is *coarser* than $\mathcal{E}$, and is obtained by *coarsening* $\mathcal{E}$. For example, the decomposition (5.14) is a refinement of (5.16) obtained by replacing the single projector $[0] + [1]$ in the latter with the two projectors [0] and [1].

According to this definition, any decomposition of the identity is its own refinement (or coarsening), and it is convenient to allow the possibility of such a *trivial* refinement (or coarsening). If the two decompositions are actually different, one is a *non-trivial* or *proper* refinement/coarsening of the other. An *ultimate* decomposition of the identity is one in which each projector projects onto a one-dimensional subspace, so no further refinement (of a non-trivial sort) is possible. Thus (5.13), (5.14), and (5.17) are ultimate decompositions, whereas (5.16) is not.

Two or more decompositions of the identity are said to be (mutually) *compatible* provided they have a *common refinement*, that is, provided there is a single decomposition $\mathcal{R}$ which is finer than each of the decompositions under consideration. When no common refinement exists the decompositions are said to be (mutually) *incompatible*. If $\mathcal{E}$ is a refinement of $\mathcal{F}$, the two are obviously compatible, because $\mathcal{E}$ is itself the common refinement.

The toy model with $M = 3$ considered in Sec. 5.2 provides various examples of compatible and incompatible decompositions of the identity. The decomposition

$$([-1] + [0]),\ \ [1] \tag{5.18}$$

is compatible with (5.16) because (5.14) is a common refinement. The decomposition

$$[-1],\ \ [p],\ \ [q], \tag{5.19}$$

where the projectors $[p]$ and $[q]$ correspond to the kets

$$|p\rangle = \big(|0\rangle + |1\rangle\big)/\sqrt{2}, \quad |q\rangle = \big(|0\rangle - |1\rangle\big)/\sqrt{2}, \tag{5.20}$$

is a refinement of (5.16), as is (5.14), so both (5.14) and (5.19) are compatible with (5.16). However, (5.14) and (5.19) are incompatible with each other: since each is an ultimate decomposition, and they are not identical, there is no common refinement. In addition, (5.19) is incompatible with (5.18), though this is not quite so obvious. As another example, the two decompositions

$$I = [x^+] + [x^-], \quad I = [z^+] + [z^-] \tag{5.21}$$

for a spin-half particle are incompatible, because each is an ultimate decomposition, and they are not identical.

If $\mathcal{E}$ and $\mathcal{F}$ are compatible, then each projector E_j can be written as a combination of projectors from the common refinement $\mathcal{R}$, and the same is true of each F_k. That is to say, the projectors $\{E_j\}$ and $\{F_k\}$ belong to the Boolean event algebra generated by $\mathcal{R}$. As all the operators in this algebra commute with each other, it follows that every projector E_j commutes with every projector F_k. Conversely, if every E_j in $\mathcal{E}$ commutes with every F_k in $\mathcal{F}$, there is a common refinement: all non-zero projectors of the form $\{E_jF_k\}$ constitute the decomposition *generated* by $\mathcal{E}$ and $\mathcal{F}$, and it is the coarsest common refinement of $\mathcal{E}$ and $\mathcal{F}$. The same argument can be extended to a larger collection of decompositions, and leads to the general rule that *decompositions of the identity are mutually compatible if and only if all the projectors belonging to all of the decompositions commute with each other.* If any pair of projectors fail to commute, the decompositions are incompatible. Using this rule it is immediately evident that the decompositions in (5.16) and (5.18) are compatible, whereas those in (5.18) and (5.19) are incompatible. The two decompositions in (5.21) are incompatible, as are any two decompositions of the identity of the form (5.13) if they correspond to two directions in space that are neither the same nor opposite to each other. Since it arises from projectors failing to commute with each other, incompatibility is a feature of the quantum world with no close analog in classical physics. Different sample spaces associated with a single classical system are always compatible, they always possess a common refinement. For example, a common refinement of two coarse grainings of a classical phase space is easily constructed using the non-empty intersections of cells taken from the two sample spaces.

As noted above in Sec. 5.2, a fundamental rule of quantum theory is that a description of a particular quantum system must be based upon a *single sample space* or decomposition of the identity. If one wants to use two or more *compatible* sample spaces, this rule can be satisfied by employing a common refinement, since its Boolean algebra will include the projectors associated with the individual spaces. On the other hand, trying to combine descriptions based upon two (or more) *incompatible* sample spaces can lead to serious mistakes. Consider, for example, the two incompatible decompositions in (5.21). Using the first, one can conclude that for a spin-half particle, either $S_x = +1/2$ or $S_x = -1/2$. Similarly, by using the second one can conclude that either $S_z = +1/2$ or else $S_z = -1/2$. However, combining these in a manner which would be perfectly correct for a classical spinning object leads to the conclusion that one of the four possibilities

$$\begin{aligned} &S_x = +1/2 \;\wedge\; S_z = +1/2, \quad S_x = +1/2 \;\wedge\; S_z = -1/2, \\ &S_x = -1/2 \;\wedge\; S_z = +1/2, \quad S_x = -1/2 \;\wedge\; S_z = -1/2 \end{aligned} \tag{5.22}$$

must be a correct description of the particle. But in fact all four possibilities are meaningless, as discussed previously in Sec. 4.6, because none of them corresponds to a subspace of the quantum Hilbert space.

5.4 Probabilities and Ensembles

Given a sample space, a probability distribution assigns a non-negative number or *probability* p_s, also written $\Pr(s)$, to each point s of the sample space in such a way that these numbers sum to 1. For example, in the case of a six-sided die, one often assigns equal probabilities to each of the six possibilities for the number of spots s; thus $p_s = 1/6$. However, this assignment is not a fundamental law of probability theory, and there exist dice for which a different set of probabilities would be more appropriate. Each compound event E in the event algebra is assigned a probability $\Pr(E)$ equal to the sum of the probabilities of the elements of the sample space which it contains. Thus "s is even" in the case of a die is assigned a probability $p_2 + p_4 + p_6$, which is $1/2$ if each p_s is $1/6$. The assignment of probabilities in the case of continuous variables, e.g., a classical phase space, can be quite a bit more complicated. However, the simpler discrete case will be quite adequate for this book; we will not need sophisticated concepts from measure theory.

Along with a formal definition, one needs an intuitive idea of the meaning of probabilities. One approach is to imagine an *ensemble*: a collection of N nominally identical systems, where N is a very large number, with each system in one of the states which make up the sample space $\mathcal{S}$, and with the fraction of members of the ensemble in state s given by the corresponding probability p_s. For example, the ensemble could be a large number of dice, each displaying a certain number of spots, with 1/6 of the members of the ensemble displaying 1 spot, 1/6 displaying 2 spots, etc. One should always think of N as such a large number that p_sN is also very large for any p_s that is greater than zero, to get around any concerns about whether the fraction of systems in state s is precisely equal to p_s. One says that the probability that a *single system* chosen at random from such an ensemble is in state s is given by p_s. Of course, any particular system will be in some definite state, but this state is not known before the system is selected from the ensemble. Thus the probability represents "partial information" about a system when its actual state is not known. For example, if the probability of some state is close to 1, one can be fairly confident, but not absolutely certain, that a system chosen at random will be in this state and not in some other state.

Rather than imagining the ensemble to be a large collection of systems, it is sometimes useful to think of it as made up of the outcomes of a large number of experiments carried out at successive times, with care being taken to ensure that these are independent in the sense that the outcome of any one experiment is not allowed to influence the outcome of later experiments. Thus instead of a large collection of dice, one can think of a single die which is rolled a large number of times. The fraction of experiments leading to the result s is then the probability p_s. The outcome of any particular experiment in the sequence is not known in advance, but a knowledge of the probabilities provides partial information.

Probability theory as a mathematical discipline does not tell one how to choose a probability distribution. Probabilities are sometimes obtained directly from experimental data. In other cases, such as the Boltzmann distribution for systems in thermal equilibrium, the probabilities express well-established physical laws. In some cases they simply represent a guess. Later we shall see how to use the dynamical laws of quantum theory to calculate various quantum probabilities. The true meaning of probabilities is a subject about which there continue to be disputes, especially among philosophers. These arguments need not concern us, for probabilities in quantum theory, when properly employed with a well-defined sample space, obey the same rules as in classical physics. Thus the situation in quantum physics is no worse (or better) than in the everyday classical world.

Conditional probabilities play a fundamental role in probabilistic reasoning and in the application of probability theory to quantum mechanics. Let A and B be two events, and suppose that $\Pr(B) > 0$. The *conditional probability of A given B* is defined to be

$$\Pr(A \mid B) = \Pr(A \wedge B)/\Pr(B), \tag{5.23}$$

where $A \wedge B$ is the event "A AND B" represented by the product AB of the classical indicators, or of the quantum projectors. Hence one can also write $\Pr(AB)$ in place of $\Pr(A \wedge B)$. The intuitive idea of a conditional probability can be expressed in the following way. Given an ensemble, consider only those members in which B occurs (is true). These comprise a *subensemble* of the original ensemble, and in this subensemble the fraction of systems with property A is given by $\Pr(A \mid B)$ rather than by $\Pr(A)$, as in the original ensemble. For example, in the case of a die, let B be the property that s is even, and A the property $s \leq 3$. Assuming equal probabilities for all outcomes,

$\Pr(A) = 1/2$. However, $\Pr(A \mid B) = 1/3$, corresponding to the fact that of the three possibilities $s = 2, 4, 6$ which constitute the compound event B, only one is less than or equal to 3.

If B is held fixed, $\Pr(A \mid B)$ as a function of its first argument A behaves like an "ordinary" probability distribution. For example, if we use s to indicate points in the sample space, the numbers $\Pr(s \mid B)$ are non-negative, and $\sum_s \Pr(s \mid B) = 1$. One can think of $\Pr(A \mid B)$ with B fixed as obtained by setting to zero the probabilities of all elements of the sample space for which B is false (does not occur), and multiplying the probabilities of those elements for which B is true by a common factor, $1/\Pr(B)$, to renormalize them, so that the probabilities of mutually exclusive sets of events sum to one. That this is a reasonable procedure is evident if one imagines an ensemble and thinks about the subensemble of cases in which B occurs. It makes no sense to define a probability conditioned on B if $\Pr(B) = 0$, as there is no way to renormalize zero probability by multiplying it by a constant in order to get something finite.

In the case of quantum systems, once an appropriate sample space has been defined the rules for manipulating probabilities are *precisely the same* as for any other ("classical") probabilities. The probabilities must be non-negative, they must sum to one, and conditional probabilities are defined in precisely the manner discussed above. Sometimes it seems as if quantum probabilities obey different rules from what one is accustomed to in classical physics. The reason is that quantum theory allows a multiplicity of sample spaces, i.e., decompositions of the identity, which are often incompatible with one another. In classical physics a single sample space is usually sufficient, and in cases in which one may want to use more than one, for example alternative coarse grainings of the phase space, the different possibilities are always compatible with each other. However, in quantum theory different samples spaces are generally incompatible with one another, so one has to learn how to choose the correct sample space needed for discussing a particular physical problem, and how to avoid carelessly combining results from incompatible sample spaces. Thus the difficulties one encounters in quantum mechanics have to do with choosing a sample space. Once the sample space has been specified, the quantum rules are the same as the classical rules.

There have been, and no doubt will continue to be, a number of proposals for introducing special "quantum probabilities" with properties which violate the usual rules of probability theory: probabilities which are negative numbers, or complex numbers, or which are not tied to a Boolean algebra of projectors, etc. Thus far, none of these proposals has proven helpful in untangling the conceptual difficulties of quantum theory. Perhaps someday the situation will change, but until then there seems to be no reason to abandon standard probability theory, a mode of reasoning which is quite well understood, both formally and intuitively, and replace it with some scheme which is deficient in one or both of these respects.

5.5 Random Variables and Physical Variables

In ordinary probability theory a *random variable* is a a real-valued function V defined everywhere on the sample space. For example, if s is the number of spots when a die is rolled, $V(s) = s$ is an example of a random variable, as is $V(s) = s^2/6$. For coin tossing, $V(\mathrm{H}) = +1/2,\ V(\mathrm{T}) = -1/2$ is an example of a random variable.

If one regards the x, p phase plane for a particle in one dimension as a sample space, then any real-valued function $V(x, p)$ is a random variable. Examples of physical interest include the position, the momentum, the kinetic energy, the potential energy and the total energy. For a particle in three dimensions the various components of angular momentum relative to some origin are also examples of random variables.

In classical mechanics the term *physical variable* is probably more descriptive than "random variable" when referring to a function defined on the phase space, and we shall use it for both classical and quantum systems. However, thinking of physical variables as random variables, that is, as functions defined on a sample space, is particularly helpful in understanding what they mean in quantum theory.

The quantum counterpart of the function V representing a physical variable in classical mechanics is a Hermitian or self-adjoint operator $V = V^\dagger$ on the Hilbert space. Thus position, energy, angular momentum, and the like all correspond to specific quantum operators. Generalizing from this, we shall think of any self-adjoint operator on the Hilbert space as representing some (not necessarily very interesting) physical variable. A quantum physical variable is often called an *observable.* While this term is not ideal, given its association with somewhat confused and contradictory ideas about quantum measurements, it is widely used in the literature, and in this book we shall employ it to refer to any quantum physical variable, that is, to any self-adjoint operator on the quantum Hilbert space, without reference to whether it could, in practice or in principle, be measured.

To see how self-adjoint operators can be thought of as random variables in the sense of probability theory, one can make use of a fact discussed in Sec. 3.7: if $V = V^\dagger$, then there is a unique decomposition of the identity $\{P_j\}$, determined by the operator V, such that, see (3.75),

$$V = \sum_j v'_j P_j, \tag{5.24}$$

where the v'_j are eigenvalues of V, and $v'_j \neq v'_k$ for $j \neq k$. Since any decomposition of the identity can be regarded as a quantum sample space, one can think of the collection $\{P_j\}$ as the "natural" sample space for the physical variable or operator V. On this sample space the operator V behaves very much like a real-valued function: to P_1 it assigns the value v'_1, to P_2 the value v'_2, and so forth. That (5.24) can be interpreted in this way is suggested by the fact that for a discrete sample space $\mathcal{S}$, an ordinary random variable V can always be written as a sum of numbers times the elementary indicators defined in (5.6),

$$V(s) = \sum_r v_r P_r(s), \tag{5.25}$$

where $v_r = V(r)$. Since quantum projectors are analogous to classical indicators, and the indicators on the right side of (5.25) are associated with the different elements of the sample space, there is an obvious and close analogy between (5.24) and (5.25).

The only possible values for a quantum observable V are the eigenvalues v'_j in (5.24) or, equivalently, the v_j in (5.32) below, just as the only possible values of a classical random variable are the v_r in (5.25). In order for a quantum system to possess the value v for the observable V, the property "$V = v$" must be true, and this means that the system is in an eigenstate of V. That is to say, the quantum system is described by a non-zero ket $|\psi\rangle$ such that

$$V|\psi\rangle = v|\psi\rangle, \tag{5.26}$$

or, more generally, by a non-zero projector Q such that

$$VQ = vQ. \tag{5.27}$$

In order for (5.27) to hold for a projector Q onto a space of dimension 2 or more, the eigenvalue v must be degenerate, and if $v = v'_j$, then

$$P_j Q = Q, \tag{5.28}$$

where P_j is the projector in (5.24) corresponding to v'_j.

Let us consider some examples, beginning with a one-dimensional harmonic oscillator. Its (total) energy corresponds to the Hamiltonian operator H, which can be written in the form

$$H = \sum_n (n + 1/2)\hbar\omega\, [\phi_n], \tag{5.29}$$

where the corresponding decomposition of the energy was introduced earlier in (5.17). The Hamiltonian can thus be thought of as a function which assigns to the projector $[\phi_n]$, or to the subspace of multiples of $|\phi_n\rangle$, the energy $(n+1/2)\hbar\omega$. In the case of a spin-half particle the operator for the z component of spin angular momentum divided by $\hbar$ is

$$S_z = +\tfrac{1}{2}[z^+] - \tfrac{1}{2}[z^-]. \tag{5.30}$$

It assigns to $[z^+]$ the value $+1/2$, and to $[z^-]$ the value $-1/2$. Next think of a toy model in which the sites are labeled by an integer m, and suppose that the distance between adjacent sites is the length b. Then the position operator will be given by

$$B = \sum_m mb\,[m]. \tag{5.31}$$

The position operator x for a "real" quantum particle in one dimension is a complicated object, and writing it in a form equivalent to (5.24) requires replacing the sum with an integral, using mathematics which is outside the scope of this book.

In all the examples considered thus far, the P_j are projectors onto one-dimensional subspaces, so they can be written as dyads, and (5.24) is equivalent to writing

$$V = \sum_j v_j |\nu_j\rangle\langle\nu_j| = \sum_j v_j [\nu_j], \tag{5.32}$$

where the eigenvalues in (5.32) are identical to those in (5.24), except that the subscript labels may be different. As discussed in Sec. 3.7, (5.24) and (5.32) will be different if one or more of the eigenvalues of V are degenerate, that is, if a particular eigenvalue occurs more than once on the right side of (5.32). For instance, the energy eigenvalues of atoms are often degenerate due to spherical symmetry, and in this case the projector P_j for the j'th energy level projects onto a space whose dimension is equal to the multiplicity (or degeneracy) of the level. When such degeneracies occur, it is possible to construct non-trivial *refinements* of the decomposition $\{P_j\}$ in the sense discussed in Sec. 5.3, by writing one or more of the P_j as a sum of two or more non-zero projectors. If $\{Q_k\}$ is such a refinement, it is obviously possible to write

$$V = \sum_k v''_k Q_k, \tag{5.33}$$

where the extra prime allows the eigenvalues in (5.33) to carry different subscripts from those in (5.24). One can again think of V as a random variable, that is a function, on the finer sample space

$\{Q_k\}$. Note that when it is possible to refine a quantum sample space in this manner, it is always possible to refine it in many different ways which are mutually incompatible. Whereas any one of these sample spaces is perfectly acceptable so far as the physical variable V is concerned, one will make mistakes if one tries to combine two or more incompatible sample spaces in order to describe a single physical system; see the comments in Sec. 5.3.

On the other hand, V cannot be defined as a physical ("random") variable on a decomposition which is *coarser* than $\{P_j\}$, since one cannot assign two different eigenvalues to the same projector or subspace. (To be sure, one might define a "coarse" version of V, but that would be a different physical variable.) Nor can V be defined as a physical or random variable on a decomposition which is incompatible with $\{P_j\}$, in the sense discussed in Sec. 5.3. It may, of course, be possible to approximate V with an operator which is a function on an alternative decomposition, but such approximations are outside the scope of the present discussion.

5.6 Averages

The average $\langle V\rangle$ of a random variable $V(s)$ on a sample space $\mathcal{S}$ is defined by the formula

$$\langle V\rangle = \sum_{s\in\mathcal{S}} p_s V(s). \tag{5.34}$$

That is, the probabilities are used to weight the values of V at the different sample points before adding them together. One way to justify the weights in (5.34) is to imagine an ensemble consisting of a very large number N of systems. If V is evaluated for each system, and the results are then added together and divided by N, the outcome will be (5.34), because the fraction of systems in the ensemble in state s is equal to p_s.

Random variables form a real linear space in the sense that if $U(s)$ and $V(s)$ are two random variables, so is the linear combination

$$uU(s) + vV(s), \tag{5.35}$$

where u and v are real numbers. The average operation $\langle\,\rangle$ defined in (5.34) is a linear functional on this space, since

$$\langle uU(s) + vV(s)\rangle = u\langle U\rangle + v\langle V\rangle. \tag{5.36}$$

Another property of $\langle\,\rangle$ is that when it is applied to a *positive* random variable $W(s) \geq 0$, the result cannot be negative:

$$\langle W\rangle \geq 0. \tag{5.37}$$

In addition, the average of the identity is 1,

$$\langle I\rangle = 1, \tag{5.38}$$

because the probabilities $\{p_s\}$ sum to 1.

The linear functional $\langle\,\rangle$ is obviously determined once the probabilities $\{p_s\}$ are given. Conversely, a functional $\langle\,\rangle$ defined on the linear space of random variables determines a unique probability distribution, since one can use averages of the elementary indicators in (5.6),

$$p_s = \langle P_s\rangle, \tag{5.39}$$

in order to define positive probabilities which sum to 1 in view of (5.9) and (5.38). In a similar way, the probability of a compound event A is equal to the average of its indicator:

$$\Pr(A) = \langle A \rangle. \tag{5.40}$$

Averages for quantum mechanical physical (random) variables follow precisely the same rules; the only differences are in notation. One starts with a sample space $\{P_j\}$ of projectors which sum to I, and a set of non-negative probabilities $\{p_j\}$ which sum to 1. A random variable on this space is a Hermitian operator which can be written in the form

$$V = \sum_j v_j P_j, \tag{5.41}$$

where the different eigenvalues appearing in the sum need not be distinct. That is, the sample space could be either the "natural" space associated with the operator V as discussed in Sec. 5.5, or some refinement. The average

$$\langle V \rangle = \sum_j p_j v_j \tag{5.42}$$

is formally equivalent to (5.34).

A probability distribution on a given sample space can only be used to calculate averages of random variables defined on this sample space; it cannot be used, at least directly, to calculate averages of random variables which are defined on some *other* sample space. While this is rather obvious in ordinary probability theory, its quantum counterpart is sometimes overlooked. In particular, the probability distribution associated with $\{P_j\}$ cannot be used to calculate the average of a self-adjoint operator S whose natural sample space is a decomposition $\{Q_k\}$ incompatible with $\{P_j\}$. Instead one must use a probability distribution for the decomposition $\{Q_k\}$.

An alternative way of writing (5.41) is the following. The positive operator

$$\rho = \sum_j p_j P_j / \mathrm{Tr}(P_j) \tag{5.43}$$

has a trace equal to 1, so it is a density matrix, as defined in Sec. 3.9. It is easy to show that

$$\langle V \rangle = \mathrm{Tr}(\rho V) \tag{5.44}$$

by applying the orthogonality conditions (5.10) to the product ρV. Note that ρ and V commute with each other. The formula (5.44) is sometimes used in situations in which ρ and V do *not* commute with each other. In such a case ρ is functioning as a pre-probability, as will be explained in Ch. 15.