qmc091.tex. Version of 5 October 2010.
Lecture Notes on Quantum Mechanics No. 9
R. B. Griffiths

Probabilities for Quantum Histories I. The Born Rule

References:

CQT = R. B. Griffiths, *Consistent Quantum Theory* (Cambridge, 2002)

Contents

---

1 Introduction

⊙ In the case of classical stochastic processes there are no absolute rules for assigning probabilities. One constructs a simplified *model* of the situation of interest (weather, stock prices, etc.), and makes a guess as to what probabilities are to be assigned; typically there are some parameters which can be adjusted on the basis of experiment.

◦ Of course people with lots of experience can make better guesses than those of us who are new to the game, but this does not mean they have a deep understanding of what is going on.

• Classical statistical mechanics is similar. The Boltzmann distribution is a guess, though there is a lot of evidence indicating that under the right circumstances (thermal equilibrium) it is a rather good guess. We do not (yet) have a systematic way of deriving it from the principles of mechanics, though there are various ways of showing that it is plausible.

⊙ the fact that classical Hamiltonian dynamics of a *closed* system, by which we mean one where the dynamical laws are given by appropriate derivatives of a (possibly time-dependent) Hamiltonian, is deterministic makes it possible to "translate" one's guess about the probability distribution for such a system at one time to something consistent with this at a later or at an earlier time.

• In particular, let us assume we know that the system is in some region $\mathcal{R}_0$ of the phase space at time t_0, and the deterministic dynamical law of time evolution maps this to a region $\mathcal{R}_1 = T(t_1, t_0)\mathcal{R}_0$ at time t_1. Then if, say, we assign at time t_0 uniform probability to the points in $\mathcal{R}_0$ and zero probability to the complement $\mathcal{R}_0^c$, then it is reasonable to do the same for $\mathcal{R}_1$ at time t_1, since T preserves phase-space "volume" (Liouville's theorem).

∘ One can also think of assigning probabilities to the *trajectories* traced out buy points in the phase space as a function of time.

⊙ If the classical system is not closed, in particular if it is a subsystem of a larger closed system, one can no longer describe the dynamics of this (sub)system by means of its Hamiltonian, so the situation is more complicated.

• Nonetheless, one can still hope in certain cases to obtain an *approximate* description of the dynamics of an open system without solving the dynamics of a larger closed system in which it is embedded, by treating the interaction between it and the rest of the world (its *environment*) by introducing some sort of random forces, i.e., by means of a suitable stochastic dynamics.

∘ A Brownian particle in a fluid driven randomly by collisions with the invisible atoms surrounding it is a well-known example of an open system.

⊙ The situation in quantum mechanics is similar, except for being a bit different. In particular, one generally needs a stochastic dynamics even for a closed quantum system.

• This requires constructing a sample space of quantum histories (discussed in a different set of notes) and then assigning probabilities to them. Once probabilities are assigned one can compute conditional probabilities and carry out statistical inference in much the same way as in classical stochastic processes.

• The Schrödinger equation (equivalently, the unitary time-development operators $T(t,t')$) can be used to compute conditional probabilities relating states of affairs at different times. To set up the full probability distribution one needs to make some additional assumption(s), such as the quantum state at a particular time, or a probability distribution at some time.

⊙ Unfortunately, the stochastic time development of closed quantum systems is not properly discussed in current textbooks. Instead, probabilities are introduced by means of *measurements*. This is an open-system approach since the measuring apparatus is part of the "environment" with which the system of interest interacts. There is nothing wrong with that, but in order to understand what is going on it is helpful to consider both the apparatus and the system it is measuring as part of a larger closed system to which the laws of quantum dynamics apply. Trying to do this when "measurement" is a necessary part of one's approach to probabilities leads to difficult conceptual problems: if the measurement apparatus is also part of a closed quantum system and one wants to say what is going on, then one needs an even bigger measurement apparatus to interact with the first one, and so on ad infinitum.

• Von Neumann was aware of this difficulty and tried to sweep it under the rug by appealing to psychology; see the final chapter VI of his book. Most physicists (and philosophers) who have studied it think that the formulation of quantum theory we owe to von Neumann contains a serious *measurement problem* that he did not resolve.

• By formulating quantum probabilities in a consistent way and paying attention to sample spaces, it is possible to avoid the measurement problem, and then describe realistic measurements in proper quantum terms.

⊙ Using unitary dynamics to assign probabilities to histories in closed quantum systems can best be discussed in several steps of increasing complexity. The *Born rule*, discussed in these notes, is sufficient for assigning probabilities to histories that involve *only two times*: an initial time and one later (though it could also be earlier) time. Section 2 gives the general formalism starting with the case of a single pure initial state and then going on to more complicated situations. Some applications of the formalism to a composite system are worked out in Sec. 3.

• Histories with three or more times require an extension of the Born rule that is not altogether straightforward, and these are considered in a later set of notes.

2 General Formalism

2.1 Single initial state

⊙ The simplest case is that of histories of the form

$$Y^k = [\psi_0] \odot [\phi_1^k], \tag{1}$$

with a single initial state $|\psi_0\rangle$, assumed to be normalized, at time t_0, and an orthonormal basis $\{\phi_1^k\}$ at time t_1. Notice that a single index k serves to label the projectors in (1). In order to construct a complete history sample space of projectors that sum to the history identity $I_0 \odot I_1$ we need another projector

$$\bar{Y} = (I_1 - [\psi_0]) \odot I_2 \tag{2}$$

to which we will assign probability zero.

⋆ Exercise. Check that the histories just described do form a sample space (mutually orthogonal projectors that sum to $\breve{I}$).

⊙ The Born rule then states that in this situation one should assign probabilities

$$\begin{aligned}
\Pr(Y^k) &= |\langle\phi_1^k|T(t_1,t_0)|\psi_0\rangle|^2 = |\langle\psi_0|T(t_0,t_1)|\phi_1^k\rangle|^2 && (3)\\
&= \langle\phi_1^k|T(t_1,t_0)\,[\psi_0]\,T(t_0,t_1)|\phi_1^k\rangle = \langle\psi_0|T(t_0,t_1)\,[\phi_1^k]\,T(t_1,t_0)|\psi_0\rangle && (4)\\
&= \mathrm{Tr}\{\,[\phi_1^k]\,T(t_1,t_0)\,[\psi_0]\,T(t_0,t_1)\} = \mathrm{Tr}\{\,[\psi_0]\,T(t_0,t_1)\,[\phi_1^k]\,T(t_1,t_0)\}. && (5)
\end{aligned}$$

(along with $\Pr(\bar{Y}) = 0$) to these different histories. All of these expressions are equivalent; sometimes one is more useful, or at least provides different insight from, another.

⋆ Exercise. Show that all these expressions are equivalent. [Hint 1. Write out the right side of (3) as the matrix element times its complex conjugate. Hint 2. $\mathrm{Tr}\{|\alpha\rangle\langle\beta|A\} = \langle\beta|A|\alpha\rangle$.]

⋆ Exercise. Show that $\sum_k \Pr(Y^k) = 1$.

∘ We did not assume that t_0 is earlier than t_1. The Born rule does not require a time ordering.

⊙ A rather common way of writing the Born rule is:

$$\Pr(Y^k) = |\langle\phi_1^k|\psi_1\rangle|^2; \quad |\psi_1\rangle := T(t_1,t_0)|\psi_0\rangle. \tag{6}$$

In words: Start with $|\psi_0\rangle$ as an initial state at time t_0. Then integrate Schrödinger's equation from t_0 to t_1 to get $|\psi_1\rangle$. Then take the absolute square of the inner product of $|\psi_1\rangle$ with $|\phi_1^k\rangle$ to obtain $\Pr(Y^k)$.

• It is immediately obvious that (6) is just another way of writing (3), so it will give the same result. It has, however, the following practical advantages. Rather than working out the entire $T(t_1,t_0)$, which could be a lot of work, one only has to find how it acts on a single ket, namely $|\psi_0\rangle$. In addition, once $|\psi_1\rangle$ has been worked out, it can be used to find inner products with what might be a rather large collection of kets forming the orthonormal basis $\{\phi_1^k\}$. And if one decides that one is more interested in some other basis $\{\bar{\phi}_1^k\}$ — well, $|\psi_1\rangle$ is already in hand, so one just needs to calculate a new set of inner products.

• A disadvantage of (6) is that one may be tempted to suppose that, in analogy with deterministic classical mechanics, $|\psi_1\rangle$ is "the state" of the quantum system at time t_1, where "state" is thought of as referring to a property. But if $|\psi_1\rangle$ is incompatible with some, or perhaps all, of the

$|\phi_1^k\rangle$, we are in difficulty, because $|\psi_1\rangle$, or more precisely the property $[\psi_1]$, is incompatible with the sample space, and we end up violating the single framework rule.

• This temptation can be resisted if one regards $|\psi_1\rangle$ (or $[\psi_1]$, which is all we really need to use in (6)) as a convenient *calculational tool* rather than some physical property of the system at time t_1. That is, it is a tool to calculate the probabilities $\Pr(Y^k)$. It helps to call this tool by a name: CQT uses "pre-probability," which means that $|\psi_1\rangle$ (or $[\psi_1]$) is being used to calculate a probability, or in fact several probabilities, whereas by itself it is neither a property nor a probability.

⋆ Exercise. Show that if at least two of the probabilities $\Pr(Y^k)$ are nonzero, then $|\psi_1\rangle$ is necessarily incompatible with at least some of the kets $\{|\phi_1^k\rangle\}$ making up the orthonormal basis.

• WARNING! The term "state" in quantum mechanics is used in a somewhat sloppy manner, and can mean different things depending on the context. Thus a pre-probability can be referred to as "the quantum state." But in the question "What is the probability that the system is in the state $|\phi_1^2\rangle$ at time t_1?" the term "state" refers to a physical property. When "the state" is a density operator the reference is (almost always) to a pre-probability. Even in classical mechanics the term "state" can have various meanings. It can refer to a point in the phase space, but in statistical mechanics it can refer to a probability distribution, as in "the Gibbs state." Sometimes a bit of sloppiness is useful, especially for informal discussion, but it can also cause confusion.

• Note that there are alternative ways to calculate the probabilities in (6) without referring to $|\psi_1\rangle$. In particular,

$$\Pr(Y^k) = |\langle\psi_0|\phi_0^k\rangle|^2; \quad |\phi_0^k\rangle := T(t_0, t_1)|\phi_1^k\rangle \tag{7}$$

That is, for each k integrate the Schrödinger equation from t_1 to t_0 to obtain $|\phi_0^k\rangle$, and then use the absolute square of its inner product with $|\psi_0\rangle$.

◦ Again, since we are assuming that $[\psi_0]$ is a property of the system at time t_0, it will not do to also include the $[\phi_0^k]$ among the properties at t_0 unless they happen to commute with $[\psi_0]$, which will in general not be the case. So regard the $|\phi_0^k\rangle$ as pre-probabilities, as calculational tools rather than as representing real quantum properties.

⊙ Since the histories $\{Y^k\}$ all have the same initial state $[\psi_0]$, the (marginal) probability of $[\psi_0]$ is 1, and therefore

$$\Pr(Y^k) = \Pr(\phi_1^k \,|\, \psi_0) \tag{8}$$

can be thought of as a *conditional probability*: the probability that at t_1 the system is in the state (has the property) $[\phi_1^k]$ given that at time t_0 it was in the state $[\psi_0]$.

⊙ Textbooks often intrepret probabilities of the form $\Pr(\phi_1^k \,|\, \psi_0)$ in terms of *measurements*: if a quantum system starts off at time t_0 in state $|\psi_0\rangle$ and a measurement is carried out at (or shortly after) time t_1 to determine which of the mutually-exclusive states $\{[\phi_1^k]\}$ it was in just before the measurement took place, the probability that the apparatus will yield the result k corresponding to $[\phi_1^k]$ is $\Pr(\phi_1^k \,|\, \psi_0)$.

• Such statements when carefully worded are correct, but tend to be confusing. What is in view is some sort of idealized measurement setup. A good measurement will reveal what was there just before the measurement took place. Naturally, "what was there" must be a sensible quantum property. Later we will discuss what what goes on in an idealized measurement.

• Some textbooks assert that the property $[\phi_1^k]$ revealed by an ideal measurement did not exist before the measurement took place; somehow it was "created" by the measurement process. This represents a serious misunderstanding (going back to von Neumann) of what measurements are all about; dispelling it requires introducing a proper quantum description of measurements, which must be deferred till a later set of notes.

∘ A related idea is that of "wave function collapse," which is discussed for a particular situation in Sec. 3.5.

2.2 Examples

⊙ As a first example consider a spin-half particle with initial state $|\psi_0\rangle = |z^+\rangle$ at time t_0, and trivial dynamics: $T(t,t') = I$. Given an orthonormal basis

$$|\phi_1^1\rangle = \alpha|z^+\rangle + \beta|z^-\rangle, \quad |\phi_1^2\rangle = -\beta^*|z^+\rangle + \alpha^*|z^-\rangle, \tag{9}$$

with $|\alpha|^2 + |\beta|^2 = 1$, the Born rule yields

$$\Pr(Y^1) = \Pr(\phi_1^1 \,|\, \psi_0) = |\alpha|^2, \quad \Pr(Y^2) = \Pr(\phi_2^2 \,|\, \psi_0) = |\beta|^2. \tag{10}$$

• If, in particular, $|\phi_1^1\rangle = |x^+\rangle$, $|\phi_1^2\rangle = |x^-\rangle$, then $|\alpha|^2 = |\beta|^2 = 1/2$, so the probability of each of the two histories $[z^+] \odot [x^+]$ and $[z^+] \odot [x^-]$ is equal to 1/2. Or, given the initial state $[z^+]$ at t_0, the conditional probability of $[x^+]$ and of $[x^-]$ at some later time t_1 is 1/2.

⊙ Next assume a less trivial dynamics: that produced by a constant magnetic field along the z axis, so that with $\Delta t = t - t'$ the time development operator $T(t,t')$ is given by

$$T(\Delta t) = \begin{pmatrix} e^{-i\omega\Delta t/2} & 0 \\ 0 & e^{i\omega\Delta t/2} \end{pmatrix}, \tag{11}$$

with ω a constant that is proportional to the strength of the magnetic field.

• Choose a family

$$[x^+] \odot \{[x^+], [x^-]\} \tag{12}$$

where the curly brackets enclose the two possibilities (the decomposition of the identity) at time t_1.

• A straightforward calculation yields

$$\Pr(x_1^+ \,|\, x_0^+) = \tfrac{1}{2}(1 + \cos\omega\Delta t), \quad \Pr(x_1^- \,|\, x_0^+) = \tfrac{1}{2}(1 - \cos\omega\Delta t). \tag{13}$$

⋆ Exercise. Carry out the calculation.

• Given that the right side of (13) depends continuously on Δt, and if we set $t_0 = 0$ then $t_1 = \Delta_t$, one is tempted to think of the probability as something like a physical property that is varying continuously with the time t_1, and if one could somehow "watch" the particle one would see this continuous variation.

∘ But such a picture is misleading, and here the textbook treatment invoking measurements has some advantage. The history family for which we have derived (13) includes only two times, t_0 and a single later time t_1, and if one were to think of an experimental test of the Born rule it would be necessaary to somehow prepare the system at t_0 in the state $[x^+]$ and then at t_1 carry out a measurement to determine if it is in $[x^+]$ or in $[x^-]$. To check a probabilistic prediction requires many measurements, and each measurement should begin by throwing away the system used previously, since a measurement can perturb a quantum system in a serious way. One has to start over with a fresh preparation of the state $[x^+]$, and again let the clock run for the desired period Δt before carrying out another measurement.

⊙ Toy alpha decay. See CQT Sec. 9.5.

2.3 Initial basis

$\odot$ The Born rule can also be used to assign probabilities to a sample space of histories of the form

$$Y^{(j,k)} = [\psi_0^j] \odot [\phi_1^k], \tag{14}$$

where $\{|\psi_0^j\rangle\}$ and $\{|\phi_1^k\rangle\}$ are any two orthonormal bases of the Hilbert space, assuming that we are considering a closed quantum system with a well-defined Hamiltonian and therefore time development operators $T(t, t')$. The two bases might be the same, but there is no requirement that this be the case.

$\odot$ Probabilities can then be assigned as in the case of a two-time Markov process, by first defining the transition matrix

$$M_{kj} = M(k, j) = |\langle\phi_1^k|T(t_1, t_0)|\psi_0^j\rangle|^2. \tag{15}$$

$\star$ Exercise. Show that $M(k, j)$ is doubly stochastic (each row and each column sums to 1).

$\circ$ Of course (15) can be written in many other ways; see (3) to (5).

• Let us use the abbreviated notation

$$\Pr(j_0, k_1) = \Pr([\psi_0^j], [\phi_1^k]), \quad \Pr(j_0) := \Pr([\psi_0^j]), \quad \Pr(k_1) := \Pr([\phi_1^k]) \tag{16}$$

for the joint distribution and the marginals; the subscript helps keep track of the time.

$\odot$ Given the transition matrix in (15) and assuming some distribution $\Pr(j_0)$ at the initial time, one can define the joint distribution in the same way as for a Markov process:

$$\Pr(j_0, k_1) = M(k, j) \Pr(j_0). \tag{17}$$

$\circ$ In particular, one might suppose that $\Pr(j_0)$ is zero for all values of j except for, say, $j = 1$, where it takes the value 1. In that case (17) gives the same result as in Sec. 2.1 if there we set $|\psi_0\rangle$ equal to $|\psi_0^1\rangle$.

• Quantum mechanics does not by itself determine $\Pr(j_0)$. The situation is analogous to classical mechanics, where to describe how a system moves as a function of time one needs not only the laws of mechanics, but also a starting state or initial condition, the choice of which is not fixed by these laws.

$\circ$ Alternatively, one could just as well assume a distribution $\Pr(k_1)$ at time t_1 and then use it to calculate the joint probability distribution

$$\Pr(j_0, k_1) = M(k, j) \Pr(k_1). \tag{18}$$

$\circ$ The Born rule does not by itself single out a direction or a "sense" of time; it does not distinguish future from past. One can assume $t_0 > t_1$ instead of $t_0 < t_1$.

2.4 Initial and final decomposition of the identity

$\odot$ It is possible to formulate the Born rule using arbitrary initial $\{P_0^j\}$ and final $\{Q_1^k\}$ decompositions of the identiy $\mathcal{H}$, in the following manner. Define a *weight* matrix

$$W(k, j) = \mathrm{Tr}\big[Q_1^k T(t_1, t_0) P_0^j T(t_0, t_1)\big] \tag{19}$$

and use it to define a conditional probability

$$\Pr(Q_1^k \,|\, P_0^j) = W(k, j)/\mathrm{Tr}(P_j) \tag{20}$$

• Using the properties of $T(t,t')$ one can show that

$$\sum_k \Pr(Q_1^k \,|\, P_0^j) = 1, \tag{21}$$

as one would expect for a conditional probability.

⋆ Exercise. Verify (21)

• One can equally well interchange the roles of t_0 and t_1, and write

$$\Pr(P_0^j \,|\, Q_1^k) = W(k,j)/\mathrm{Tr}(Q_k). \tag{22}$$

• Once again, the Born rule does not single out a direction of time; the scheme for assigning probabilities is *time reversible.*

◦ Do not confuse this with *time reversal symmetry*, often denoted by T, as in the PCT theorem. This T is a (possible) symmetry of the Hamiltonian H. Whether or not H has this symmetry it is Hermitian, so the corresponding time-development operators are unitary, and that unitarity is the basis of the time reversibility of the Born rule as discussed above.

◦ The time-reversibility of the Born rule is unfortunately somewhat obscured in the textbook approach which emphasizes measurements, because real measurements are irreversible in the thermodynamic sense. That is a separate issue.

3 Composite Systems

3.1 Introduction

• Consider a bipartite system with Hilbert space $\mathcal{H}_a \otimes \mathcal{H}_b$. The Born rule is formally just the same as for a single system, but the fact that the probabilities at time t_1 refer to a composite system introduces some new concepts and terminology.

• To keep the discussion simple, we will assume an initial pure state $|\psi_0\rangle$ which evolves under unitary time evolution to $|\psi_1\rangle$, and discuss probabilities at time t_1 using $|\psi_1\rangle$ as the pre-probability. All probabilities will be conditional on $|\psi_0\rangle$, so will be denoted by $\Pr(\cdots)$ instead of the clumsier $\Pr(\cdots \,|\, \psi_0)$, and since only properties at t_1 are under consideration, in this section we drop the subscript 1 from various operators and labels.

⊙ Let us suppose that we are interested in a decomposition of the identity $\{P^j\}$ for system a and another decomposition $\{Q^k\}$ for system b, which could correspond in either case (or both cases) to an orthonormal basis. Thus we shall want to consider probabilities of the form

$$\Pr(P^j, Q^k) = \langle\psi_1|P^j \otimes Q^k|\psi_1\rangle \tag{23}$$

with marginals

$$\Pr(P^j) = \sum_k \Pr(P^j, Q^k) = \langle\psi_1|P^j \otimes I_b|\psi_1\rangle, \quad \Pr(Q^k) = \langle\psi_1|I_a \otimes Q^k|\psi_1\rangle. \tag{24}$$

◦ Note that $P^j \otimes I_b$ is often abbreviated to P^j

3.2 Marginals and reduced density operators

⊙ If one is interested in one of the marginals, say $\Pr(P^j)$, and not in the entire distribution (23) it is sometimes convenient to write it in the form

$$\Pr(P^j) = \mathrm{Tr}(\rho_a P^j) \tag{25}$$

where the *reduced density operator* ρ_a on $\mathcal{H}_a$ can be defined as a partial trace

$$\rho_a = \mathrm{Tr}_b[\psi_1] \tag{26}$$

over the projector $[\psi_1] = |\psi_1\rangle\langle\psi_1|$, which is an operator on $\mathcal{H}_a \otimes \mathcal{H}_b$.

• There is a great deal of lore, not all of it helpful, associated with density operators in quantum textbooks and the literature of quantum foundations. For present purposes let us take the attitude that ρ_a is simply a convenient calculational tool for obtaining probabilities which can also be written without reference to it, as in (24). Thus it is an example, as is $|\psi_1\rangle$, of a pre-probability.

• Why is it convenient? Suppose that $d_a = 2$ and $d_b = 50$. Then ρ_a is a small 2×2 matrix, easy to parametrize, that can be used for various different decompositions $\{P^j\}$, whereas $|\psi_1\rangle$ or the corresponding $[\psi_1]$ might be quite a mess.

• Convenience, however, also means there are limitations on what one can calculate from ρ_a *by itself*. In particular, it provides no information about the *correlations* that may be present between system a and b. If one is interested in these, some other mathematical tool is needed.

◦ Compare with ordinary probability theory. The joint distibution of two random variables $\Pr(V = v, W = w)$ contains more information than $\Pr(V = v)$ or (except when they are independent) both $\Pr(V = v)$ and $\Pr(W = w)$.

3.3 Correlations and conditional states

⊙ Suppose we are interested in correlations between a and b given the pre-probability $|\psi_1\rangle$. The simplest situation is that in which one of the decompositions corresponds to an orthonormal basis; let us suppose that $P^j = |a^j\rangle\langle a^j|$. One can then consider the expansion

$$|\psi_1\rangle = \sum_j |a^j\rangle \otimes |\beta^j\rangle, \tag{27}$$

where the $|\beta^j\rangle$ are simple expansion coefficients: not normalized nor orthogonal and perhaps not even a basis of $\mathcal{H}_b$. But useful in that we can write the joint distribution in the form:

$$\Pr(P^j, Q^k) = \Pr(a^j, Q^k) = \langle\beta^j|Q^k|\beta^j\rangle. \tag{28}$$

Another useful thing to note is that

$$\Pr(P^j) = \Pr(a^j) = \langle\beta^j|\beta^j\rangle. \tag{29}$$

⋆ Exercise. Derive (28) starting with (23), and verify (29).

• By combining (28) and (29) one arrives at a formula for the conditional probability:

$$\Pr(Q^k \mid P^j) = \Pr(Q^k \mid a^j) = \langle\bar\beta^j|Q^k|\bar\beta^j\rangle, \quad |\bar\beta^j\rangle := |\beta^j\rangle/\|\beta^j\|, \quad \|\beta^j\| = \sqrt{\langle\beta^j|\beta^j\rangle}. \tag{30}$$

◦ In words: one carries out the expansion in (27), but then renormalizes the ket $|\beta^j\rangle$ before using it as a pre-probability to calculate the conditional probability of Q^k given a^j.

• One can call $|\beta^j\rangle$ or the normalized $|\bar\beta^j\rangle$ a *conditional ket* or *conditional state*: it is a pre-probability, since it is used to compute a (conditional) probability distribution.

3.4 Example: singlet-state correlations

⊙ The following example will illustrate the use of conditional states. Suppose that a and b are two spin-half particles, and

$$|\psi_1\rangle = (1/\sqrt{2})(|z_a^+ z_b^-\rangle - |z_a^- z_b^+\rangle), \tag{31}$$

the spin singlet state. (It corresponds to zero total angular momentum of the the two particles.)

• We use the z basis for a: $P^1 = [z_a^+]$, $P^2 = [z_a^-]$. The conditional states of b are then:

$$|\beta^1\rangle = |z^-\rangle/\sqrt{2}, \quad |\bar{\beta}^1\rangle = |z^-\rangle; \quad |\beta^2\rangle = |z^+\rangle/\sqrt{2}, \quad |\bar{\beta}^2\rangle = |z^+\rangle. \tag{32}$$

∘ By squaring the factor of $1/\sqrt{2}$ we obtain the marginal probabilities $\Pr(z_a^+) = \Pr(z_a^-) = 1/2$.

• Let us choose for b a decomposition $Q^1 = [z_b^+]$, $Q^2 = [z_b^-]$. Then by means of (30) we find that:

$$\Pr(z_b^+ \,|\, z_a^+) = 0, \quad \Pr(z_b^- \,|\, z_a^+) = 1, \quad \Pr(z_b^+ \,|\, z_a^-) = 1, \quad \Pr(z_b^- \,|\, z_a^+) = 0. \tag{33}$$

• If we choose a different decomposition, say $Q^1 = [x_b^+]$, $Q^2 = [x_b^-]$, then of course we get a different answer:

$$\Pr(x_b^+ \,|\, z_a^+) = \Pr(x_b^- \,|\, z_a^+) = 1/2, \quad \Pr(x_b^+ \,|\, z_a^-) = \Pr(x_b^- \,|\, z_a^+) = 1/2. \tag{34}$$

∘ Needless to say, one should not try and combine the incompatible (in the quantum sense) results in (33) and (34): We cannot conclude from z_a^+ that for particle b it is the case that $S_{bz} = -1/2$ and at the same time S_{bx} has equal probabilities to take the values $+1/2$ and $-1/2$.

⋆ Exercise. Work out the conditional probabilities in the case in which $Q^1 = [w_b^+]$, $Q^2 = [w_b^-]$ for w some arbitary direction in space corresponding to polar coordinates θ, ϕ.

⊙ We could also use a different decomposition for particle a. Let us consider the general case of $P^1 = [w_a^+]$, $P^2 = [w_a^-]$, where w denotes an arbitary direction in space (arbitary point on the Bloch sphere). We then need to carry out the expansion (27) in the w basis. It turns out that up to a phase (which depends upon the phase conventions one uses for spin-half states, but has no effect on the physics) one can write

$$|\psi_1\rangle = (1/\sqrt{2})(|w_a^+ w_b^-\rangle - |w_a^- w_b^+\rangle). \tag{35}$$

• As a consequence the conditional kets are now

$$|\beta^1\rangle = |w^-\rangle/\sqrt{2}, \quad |\bar{\beta}^1\rangle = |w^-\rangle; \quad |\beta^2\rangle = |w^+\rangle/\sqrt{2}, \quad |\bar{\beta}^2\rangle = |w^+\rangle, \tag{36}$$

and extracting conditional probabilities for a given decomposition for particle b proceeds in an obvious way.

3.5 Wave function collapse

⊙ While (30) is a sensible, and often an efficient, way of computing conditional probabilities, it is usually presented in textbooks under the heading of "wave function collapse," or similar words, and described as something that happens when a *measurement* is carried out. This has unfortunately given rise to much nonsense, including the notion that measurements on quantum systems produce effects at very great distances (thinking of a and b as a long distance apart) at speeds which can exceed the speed of light, which seems to bring quantum mechanics into conflict with relativity theory.

• A discussion of what measurements actually measure when described in proper quantum terms must await the introduction of additional material (on multitime histories), but the following classical analogy should help in avoiding the pitfalls associated with wave function collapse.

⊙ Charlie in Chicago takes two slips of paper, one red and one green, places them in two opaque envelopes, and after shuffling them so that he himself does not know which is which, addresses one to Alice in Atlanta and the other to Bob in Boston, both of whom know the protocol Charlie is following. Upon receipt of the envelope addressed to her Alice opens it and sees a red slip of paper. From this she can immediately conclude that the slip in Bob's envelope is green, whether or not Bob has already opened his envelope, or will ever do so. Her conclusion is not based on a belief that opening her envelope to "measure" the color of the slip of paper has any influence on the color of Bob's slip. Instead she is employing statistical reasoning in the following way.

• Before Alice opens the envelope she (or Bob or Charlie) can assign probabilities to the various situations as follows:

$$\Pr(R_a, R_b) = \Pr(G_a, G_b) = 0, \quad \Pr(R_a, G_b) = \Pr(G_a, R_b) = 1/2, \tag{37}$$

where R_a means a red slip in Alice's envelope, G_b a green slip in Bob's, etc., and Pr() in (37) is the joint probability distribution of the colors. From the usual rule for conditional probabilities it follows that

$$\Pr(R_b \,|\, R_a) = 0, \quad \Pr(G_b \,|\, R_a) = 1, \tag{38}$$

and this is the conditional probability distribution that Alice uses to infer the color of Bob's slip knowing that the one in her envelope is red. One might say that she uses the outcome of her observation to "collapse" the initial (marginal) probability distribution $\Pr(R_b) = \Pr(G_b) = 1/2$ corresponding to (37) onto the conditional probability distribution (38). But the colors of the two slips of paper are not at all affected by Alice's "measurement." It is her *knowledge* of the world that changes, in a way which we do not find at all surprising. The "collapse," if one calls it that, refers to a method of reasoning, not a physical effect.

⊙ To relate this analogy to the previous example, think of $[z^+]$ and $[z^-]$ as analogous to the properties R and G, respectively. Assume that a measurement is carried out on particle a to determine the value of S_{az}. If it is a good measurement it reveals a property of particle a before the measurement was carried out, and using conditional probabilities obtained from conditional kets, or by working out the joint probability distribution of S_{az} and S_{bz} using (23), with $|\psi_1\rangle$ from (31), one arrives at the set of conditional probabilities in (33) which are the analogs of (37). The probabilistic reasoning required when a and b are spin-half particles is the same as when they are colored slips of paper.

• What brings about additional structure in the spin-half case is that it is rather natural to think of using a different, incompatible sample space, say $\bar{P}^1 = [w_a^+]$, $\bar{P}^2 = [w_a^-]$, where w is neither z nor $-z$, because one can easily imagine measuring S_{aw} rather than S_{az}. However, simultaneous measurements of S_{aw} and an S_{az} are not possible (as correctly stated in the textbooks). The component which Alice (now thought of as in a physics laboratory) chooses to measure determines what she learns about particle a, and therefore what she can say about particle b. But this choice has absolutely no influence on particle b, as one can demonstrate by carrying out the requisite calculations in detail. This is done (in a somewhat tedious way) in Ch. 23 of CQT.