10-18-2011

Why Probability Appears in Quantum Mechaincs

Jerome Blackman
Syracuse University

Wu Teh Hsiang
Syracuse University

Why Probability Appears in Quantum Theory

Jerome Blackman, Syracuse University, Dept. of Mathematics, Emeritus
Wu Teh Hsiang, Syracuse University, Dept. of Mathematics

1. Summary: Early in the development of quantum theory Bohr introduced what came to be called the Copenhagen interpretation. Specifically, the square of the absolute value of the wave function was to be used as a probability density. There followed lengthy arguments about this ranging from alternative universes to Schrödinger's cat. Einstein famously remarked "I am convinced that He (God) does not play dice."

The purpose of this paper is to present a mathematical model of the measuring process that shows that the Copenhagen interpretation can actually follow from the fact that the time development of quantum systems is governed by the usual one parameter group of unitary transformations e^{-iHT} and that probability enters into the theory in the way it usually does in physics, namely, by having a large number of deterministic equations that can only be handled probabilistically.

In the literature on the measuring problem various arguments are presented to show that the above outlined plan will not work so we will start with some comments on the errors in those arguments.

2. Introduction: The arguments on the impossibility of deriving the Copenhagen interpretation from the Schrödinger equation generally form into two classes. The first is the probability argument which points out that the Schrödinger equation is a deterministic equation and has no room for a probabilistic interpretation. We, on the other hand, start with the basic idea that any quantum measurement involves macroscopic equipment. Since this equipment is not at absolute zero it is subject to small but important variations, not only in time during the measurement, but also from one measurement to another. It is this variation and the entanglement of the particle with the measuring equipment that leads to the different results of repeated measurements. It is interesting that in this model it is the influence of the particle on the measuring equipment that, through the entanglement process, has a main role. By 'measuring equipment' we mean not only the final device on which a number is read but rather the entire experimental machinery.

The second argument is based on linearity. If a particle is measured in any eigenstate of the observable being measured the result is the same eigenstate. By linearity the Schrödinger equation predicts the result for an

initial state that is a linear combination of the two should be a linear combination contrary to experimental evidence. The error here is that the linearity refers to states in the space of the particle and the measuring device and not that of the particle alone. The same error accounts for arguments involving the fact that no matter how complicated the scheme involving the Schrödinger equation the transformation from the initial state to the final should be a unitary transformation and not a projection as actually seems to happens.

If we are faced with the possibility of a large number of possible configurations of the measuring device and the possibility that even if we start with a given initial configuration it may vary in many different ways during the course of the experiment, the most convenient way to treat the problem is to suppose we have a Hilbert space $\mathcal{H}_m$ as the configuration space for the measuring device and that $\mathcal{H}_m$ is large enough to contain all 'permissible' states where 'permissible' will be defined below. If the initial state of the measuring device to the finest detail at the particle level and the initial state of the particle are known there is, in principle, a Hamiltonian that describes the future state of the system. Since the initial state of the particle will be assumed to be described as a linear combination of the eigenstates of an observable we can assume that a state of the system is determined by the initial state of the measuring system.

Let Λ be an index set so that each $\lambda \in \Lambda$ represents a permissible state of the measuring equipment. Let $\mathcal{H}_p$ be the Hilbert space of the particle so that the tensor product $\mathcal{H}_m \otimes \mathcal{H}_p = \mathcal{H}$ is the Hilbert space of the interacting particle and measuring device together. Let $\boldsymbol{P}$ be an observable of the particle with a complete set of orthonormal eigenvectors $|p_i\rangle$ and corresponding eigenvalues p_i. Let $|q_j\rangle$ be an orthonormal basis for $\mathcal{H}_m$. We write these as denumerable sets and all additions as sums for simplicity of notation with the understanding that there would be no difficulty in replacing the sums with appropriate integrals. In $\mathcal{H}$ the Hamiltonians will be designated by H_λ with λ indicating the particular Hamiltonian which is associated with $|g_\lambda\rangle$ the associated initial state of the equipment. We will assume non-degeneracy in the operator $\boldsymbol{P}$ so that each eigenstate has a unique eigenvector except for the usual ambiguity of phase. This is not a serious restriction but it does avoid unimportant complexity,

The set $\{|q_j\rangle \otimes |p_i\rangle\}$ is a basis for $\mathcal{H}$ and the time development of the system is governed by a one parameter group of unitary transformations such as described by the Schrödinger or Dirac equations. The action of this

group in the time interval $[t_1, t_2]$ is indicated by the operator $U([t_1, t_2], H_\lambda)$. The main property of this group that we will need is precisely the group property

(1) $$U([t_0, t_2], H_\lambda) = U([t_1, t_2], H_\lambda)\, U([t_0, t_1], H_\lambda) \quad \text{for } t_0 \le t_1 \le t_2 .$$

We assume that the measurement occurs in the interval [0, T] and, for simplicity, write

$$U([0, t], H_\lambda) = U_\lambda(t).$$

If the initial state is

(2) $$\sum_j b_\lambda^j \left|q_j\right\rangle \otimes \sum_i a^i \left|p_i\right\rangle$$

then at time t it is of the form

(3) $$U_\lambda(t) \sum_j \sum_i b_\lambda^j a^i \left|q_j\right\rangle \otimes \left|p_i\right\rangle = \sum_{i,j} a_\lambda^{i,j}(t) \left|q_j\right\rangle \otimes \left|p_i\right\rangle \text{ where } a_\lambda^{i,j}(0) = b_\lambda^j a^i.$$

For each $\lambda \in \Lambda$ this describes a curve in the unit sphere of $\mathcal{H}$ which starts at (2), and the burden of the Copenhagen interpretation is that at t = T it ends at a point $\left|g_\lambda\right\rangle \otimes \left|p_i\right\rangle$ with probability $\left|a^i\right|^2$. Here $\left|g_\lambda\right\rangle$ is some state vector for the measuring apparatus which may depend on λ. In order for this to make any sense it is necessary to introduce a probability measure into the set Λ. Before doing that we need to decide, in the formalism above, exactly what is the requirement of a $U_\lambda(t)$ to be a valid measurement of the observable P.

3. What is a measurement? Feynman and Hibbs (2, p22) asked "Why can we only predict the probability that a given experiment can lead to a certain result? Almost certainly it arises from the need to amplify the effect of a simple atomic event to such a level that they may be readily observed by large systems."

Starting from this observation it would seem that there are two essential steps in measuring the value of an observable. The first is to find a reaction that magnifies in some way the quantum level to a macroscopic level. The second is to do it in such a way that the value to be measured is not changed in the process. This seems to be done by using a process that would not change the value such as energy or momentum if one were dealing with a classical particle. The problem is to formulate this requirement in terns that are appropriate to the formalism of quantum theory and that takes into account that the particle is generally in a superposition of states. To do this let ψ be a subset of the set of eigenvectors of the eigenvectors of P and $\mathcal{H}_{p,\psi}$ the complex Hilbert space with the elements of ψ as basis. The spaces $\mathcal{H}_m \otimes \mathcal{H}_{p,\psi}$ are all subspaces of $\mathcal{H}_m \otimes \mathcal{H}_p = \mathcal{H}$.

The condition for a permissible state can now be formulated in terms of the Hamiltonian. It consists of two parts.

(4) (a) The experiment must be designed so that the final state of the measuring equipment indicates the final state of the particle.

(b) The following equation must hold:

$H(\mathcal{H}_m \otimes \mathcal{H}_{p,\psi}) \subset (\mathcal{H}_m \otimes \mathcal{H}_{p,\psi})$. This need be true only for those subsets ψ of the eigenvalues which are in the range of the experiment. In other words, if the energy of a laboratory generated electron is under consideration it is not necessary to allow for energy in the range of cosmic rays. Since iH is the generator of the unitary group of transformations (4) states that if at any time t_0 the right side of equation (3) only contains the set ψ of eigenstates it will contain no more at any subsequent time.

If the initial state of the measuring device and particle is

(5) $$|g_\lambda\rangle \otimes \sum_i a^i |p_i\rangle$$

then the path in the unit sphere is, from (4),

(6) $$\sum_i a^i |g_{\lambda,i}(t)\rangle \otimes |p_i\rangle .$$

At this point assume the $\{|p_i\rangle\}$ are a finite set with i taking the values 1, 2. …n. The more general case will be treated below. Remembering that the $\{a^i |g_{\lambda,i}(t)\rangle\}$ are elements of $\mathcal{H}_m$ it is convenient to think of them as generalized coefficients of an n dimensional vector. As the curve (6) develops in time the number of non- zero terms in (6) is reduced by 1 as soon as one of the generalized coefficients has value 0. This is a consequence of (4b). It is now possible to see why at the end of the measurement it may be reduced to one term with the particle in an eigenstate of ***P***.

Because we are only interested in the end points of the curves (6) it is convenient to map them into curves in R_n . Let e_1 = (1, 0… 0), e_2 = (0, 1… 0) … e_n = (0, 0 …0, 1). The smallest convex set containing these n points is a simplex denoted by S_{n-1}. S_1 is a line segment, S_2 is a triangle, S_3 is a tetrahedron and so on. Observe that S_{n-1} consists of an n-1 dimensional subset of R_n plus boundary sets of lower dimensions. S_{n-1} can also be characterized as the set of all points $(f_1, f_{2,}...f_n)$ such that $\sum_1^n f_i = 1$ with $f_i \geq 0$. Let $|a^i |g_{\lambda,i}(t)\rangle|^2 = c^i_\lambda(t)$. Then, since the curves (6) lie in the unit sphere

$\sum_1^n c_\lambda^i(t) = 1$, and, obviously, the $c^i(t)$ are ≥ 0. Let ϕ be the map which takes (6) into $(c_\lambda^1(t), c_\lambda^2(t) \ldots c_\lambda^n(t))$. Then ϕ maps the curves (6) onto curves in S_{n-1}. It is easy to see that at t = 0 the curve starts at $(c_\lambda^1(0), c_\lambda^2(0) \ldots c_\lambda^n(0))$ and that $c_\lambda^k(0) = |a^k|^2$ for all λ. Set $c_\lambda^k(0) = c_\lambda^k$. The Copenhagen interpretation, reformulated for the curves in S_{n-1}, states that if the curve starts at $(c_\lambda^1, c_\lambda^2, \ldots c_\lambda^n)$ it will with probability c_λ^k end at the point e_k.

If we only consider points $(f_1, f_2, \ldots f_n)$ in S_{n-1} where k of the coordinates are zero then we get sub simplexes of dimension n – k -1. There are $\binom{n}{k}$ of them where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ if $n \geq k$ and $\binom{n}{k} = 0$ if $n < k$.

It is clear that the curves we are considering have the property that every time they have a component vanish they enter a sub simplex and they stay in that sub simplex for all subsequent times.

The case of an infinite spectrum can be reduced to the finite spectrum case by recognizing that any measurement has only a finite accuracy. This implies that the spectrum can be covered by a finite set of non overlapping intervals so that all points in an interval will yield the same measurement. Let the points of the spectrum that are in the kth interval be indicated by p_{k1}, p_{k2}… and suppose there are n intervals. Then the expression (6) can be written as

(6a) $$\sum_{k=1}^n \sum_i a^{i,k} \left| g_{\lambda,i,k} \right\rangle \otimes \left| p_{i,k} \right\rangle.$$

Now let

$$c_\lambda^k(t) = \sum_i |a^{i,k} \left| g_{\lambda,i,k}(t) \right\rangle|^2 \quad \text{for k = 1,…n.}$$

This defines a map φ' of the paths in the unit sphere of $\mathcal{H}$ into S_{n-1} and $c_\lambda^k(t) = 0$ if and only if $a^{i,k} \left| g_{\lambda,i,k}(t) \right\rangle = 0$ for all i. Moreover in this case we can weaken the condition (4b). As it stands (4b) represents a finite number of restrictions in the finite spectrum case but an infinite number if the spectrum is infinite. But we can redefine ψ to be any subset of the integers 1, 2,…n and $\mathcal{H}_{p,\psi}$ to be the complex Hilbert space whose basis consist of all the eigenvectors whose eigenvalues lie in any of intervals indexed by ψ. This leaves a finite number of conditions and only with the problem of satisfying (4b) to within the accuracy of his measurements. Of course (4a)

must be interpreted to mean the final state of the measuring equipment must indicate the interval in which the final eigenvalues lies.

It is interesting to see what happens to the linearity argument against our approach to the measurement problem. According to the work above if the initial state of the system is initially in the states $|g_{\lambda,i}\rangle \otimes |p_i\rangle$ for i = 1 ,2 the final states will be $|g'_{\lambda,i}\rangle \otimes |p_i\rangle$ where by (4a) the $|g'_{\lambda,i}\rangle$ indicate the final states of the measuring equipment and must each indicate the eigenvectors $|p_1\rangle$ and $|p_2\rangle$ respectively. If we start with a linear combination of the two initial states linearity gives $a|g'_{\lambda,1}\rangle \otimes |p_1\rangle + b|g'_{\lambda,2}\rangle \otimes |p_2\rangle$. But this is still a state in which the particle and the measuring equipment are intertwined and indicates the measuring process is not complete.

We can now turn to the problem of introducing a probability measure on the set of paths in S_{n-1}. Although we will have no need to do so this measure can be pulled back to the original paths in the unit sphere of $\mathcal{H}$ via the inverse map ϕ^{-1} along with the appropriate Borel field of sets.

4. Probability. One of the problems is that it is difficult to handle the case where the time interval is continuous as is the set S_{n-1}. We will follow the not very original program of using discrete approximations, showing that the desired result holds no matter how fine the approximation. Choose an arbitrary integer m and let Δ_{n-1} be the set of points $(\frac{k_1}{m}, \frac{k_2}{m}, \ldots \frac{k_n}{m})$ where all the k_i are ≥ 0 and $\sum_{i=1}^{n} k_i = m$. Then $\Delta_{n-1} \subset S_{n-1}$. We also divide the time interval by an arbitrarily large set of equidistant points

$$0 = t_0 < t_1 < t_2 < \ldots < t_r = T .$$

We can think of the points of Δ_{n-1} as being states of the system even though each one represents many quantum states in $\mathcal{H}$. This is not important because the Copenhagen interpretation only refers to the final state and does not make any statement about the phase of the final state as long as it is an eigenstate of P.

We have already introduced notation for n of these states namely e_1 to e_n. Label the states by s_i with the convention that the first n of these are the e_i for i = 1, …n. A path is just a sequence of these states and we can assign probability to paths by assigning transition probabilities. Because the original paths in the Hilbert space after t_0 depends only on the state at that time it is reasonable to use a Markov chain as a model and not unreasonable

to assume it is stationary. Let $P(s_0, s_1)$ be the probability of a transition from s_0 to s_1 in one step. Unfortunately the problem of assigning values to $P(s_0, s_1)$ is complicated by the need to take into account the fact that the paths cannot leave a sub simplex after entering it. We now have to introduce some notation but the structure will be clearer if one keeps in mind the simple case of defining a stationary Markov process on the integral valued points in the plane. In that case we would have only to define the probability of moving up, down, left or right to get a transition probability. The added complexity is all due to the higher dimension and the sub simplex restriction where we have to assign probability 0 to leaving a sub simplex.
Let $s_j = (a_{j,1}, a_{j,2} \ldots a_{j,n})$.

Definition 1. Let $|s_1 - s_2| = \sum_{j=1}^{n} (a_{1,j} - a_{2,j})^2$.

Definition 2. Let $l(s_0) = \{j | \in [1, 2, \ldots n] \mid a_{0.j} > 0\}$.

Definition 3. Let $i(s_0) = \# l(s_0)$ where #(A) is the cardinality of the set A. $i(s_0)$ is the index of s_0 and $i(s_0) - 1$ is the dimension of the sub simplex in which s_0 lies.

Definition 4. $N(s_0) = \{ s \in \Delta_{n-1} | \, |s - s_0| \leq \frac{2}{m^2}, l(s) \subseteq l(s_0)\}$ is the neighborhood of s_0.

In the last definition, the condition $|s - s_0| \leq \frac{2}{m^2}$ means that the coordinates of s and s_0 can differ in either 0 or 2 places and in the latter case by only $\frac{1}{m}$ in each place. The condition $l(s) \subseteq l(s_0)\}$ means that a point in the neighborhood of s_0 must lie in the same face as does s_0.

$$N_{j-}(s_0) = \{s \in N(s_0) \mid a_j - a_0 = \frac{-1}{m}\}$$

Definition 5. $N_j(s_0) = \{s \in N(s_0) \mid a_j - a_0 = 0\}$ $\qquad s = (a_i, \ldots a_n)$

$$N_{j+}(s_0) = \{s \in N(s_0) \mid a_j - a_0 = \frac{1}{m}\}$$

With a bit of counting we find that

$$\#N(s_0) = 1 + 2\binom{i}{2}$$

$$\#N_j(s_0) = 1 + 2\binom{i-1}{2}$$

$$\#N_{j-}(s_0) = \#N_{j+}(s_0) = i - 1 \text{ and}$$

$$\#N(s_0) = \#N_{j-}(s_0) + \#N_j(s_0) + \#N_{j+}(s_0).$$

The formal statement of the properties of the transition matrix $P = (P(s_i, s_j))$ which we require are as follows:

1 $P(s_i, s_j) = 0$ if $s_j \notin N(s_0)$

2. $\sum_{s_1 \in \Delta_{n-1}} P(s_0, s_1) = \sum_{s_1 \in N(s_0)} P(s_0, s_1) = 1$

3. $\sum_{s_1 \in N_{j+}(s_0)} P(s_0, s_1) = \sum_{s_1 \in N_{j-}(s_0)} P(s_0, s_1)$

4. $P(s_0, s_0) < 1$ if $s_0 \neq e_j$ for some j, $1 \leq j \leq n$

Condition 1 is the statement that only transitions to neighboring states are allowed. Since the original curves in the Hilbert space were continuous this is not unreasonable. Also, since the states e_i have only themselves in their neighborhood these states are absorbing, i.e. once in one of these the path remains there. Condition 2 is just the statement that the transition matrix is stochastic. Condition 3 is the symmetry condition that forbids a preferred drift to a particular e_k. Condition 4 prevents the process from stopping except in one of the states e_k. All together these conditions allow a large number of admissible stochastic matrices.

Let P represent the transition matrix $P(s_i, s_j)$. Then P^r is the transition matrix whose entries are the probabilities of a transition form the state s_i to the state s_j in r steps. Fortunately matrices which satisfy conditions 1 to 4 have been thoroughly investigated ((1), chapter 5). Under these conditions the unique limit of as $r \to \infty$ exists. The convergence is exponentially fast, and, with probability 1 the transition states (the non-absorbing states) are occupied for only a finite number of steps. Furthermore, if $\lim_{r \to \infty} P^r = S$ then S is the unique solution of the equation PS = S. While there are only a finite number of steps in any partition of the interval (0,T) we assume that T is large enough to allow for the finite number of steps referred to in the preceding sentence. It follows that all that remains is to show that the S

predicted by the Copenhagen interpretation does indeed satisfy this equation. If the particle start initially in the state $s_0 = \left(\frac{k_1}{m}, \frac{k_2}{m}, \ldots \frac{k_i}{m}, \ldots \frac{k_n}{m}\right)$ the prediction is that the final state is e_i with probability $\frac{k_i}{m}$. Therefore the prediction for the form of $\overline{\overline{P}}$ given by the Copenhagen interpretation is

$$S(s_0, e_i) = \frac{k_i}{m} \text{ and}$$
$$S(s_0, s_i) = 0 \text{ if } s_i \notin (e_1, e_2, \ldots e_n).$$

5. Statement and proof of the main theorem.
Theorem: S = PS.
Proof: From the definition of S it is immediate that for each $s_0 \in \Delta_{n-1}$

(7) $$S(s_0, s_i) = 0 \text{ if } s_i \notin (e_1, e_2, \ldots e_n).$$

Matrix multiplication yields $(PS)(s_0, s_i) = \sum_{s \in \Delta_{n-1}} P(s_0, s) S(s, s_i)$. By equation (7) if $s_i \notin (e_1, e_2, \ldots e_n)$ then the second term in this sum is 0 so $PS(s_0, s_i) = 0$.

If $s_i = e_j$ for some j, $1 \le j \le n$ then

$$(PS)(s_0, e_j) = \sum_{s \in \Delta_{n-1}} P(s_0, s) S(s, e_j) = \sum_{s \in N(s_0)} P(s_0, s) S(s, e_j).$$

We distinguish two cases.

(a) $j \notin l(s_0)$. This implies that $j \notin l(s) \ \forall s \in N(s_0)$. Hence $(PS)(s_0, e_j) = 0$.

(b) $j \in l(s_0)$. $(PS)(s_0, e_j) = \sum_{s \in N(s_0)} P(s_0, s) \sum_{s \in N(s_0)} P(s_0, s) S(s, e_j)$

$$= \{ \sum_{s \in N_j(s_0)} + \sum_{s \in N_{j+}(s_0)} + \sum_{s \in N_{j-}(s_0)} \} P(s_0, s) S(s, e_j)$$

$$= \frac{k_j}{m} \sum_{s \in N_j(s_0)} P(s_0, s) + \frac{k_j + 1}{m} \sum_{s \in N_{j+}(s_0)} P(s_0, s) + \frac{k_j - 1}{m} \sum_{s \in N_{j-}(s_0)} P(s_0, s)$$

$$= \frac{k_j}{m} [\sum_{s \in N(s_0)} P(s_0, s)] + \frac{1}{m} [\sum_{s \in N_{j+}} P(s_0, s) - \sum_{s \in N_{j-}} P(s_0, s)]$$

$$= \frac{k_j}{m} \text{ because of the symmetry condition in the definition of P.}$$

Hence $PS(s_0, s) = S(s_0, s) \ \forall \ s_0, s \in \Delta_{n-1}$, i.e. PS = S and the theorem is proved.

(6)Conclusion. How much the above describes the actual physical process of measurement seems to depend on two factors. The first is the question of whether the condition (4b) is a realistic description of what the

experimenter is doing when a quantum measurement is performed. The second is the question of whether the measure introduced into the space of paths is realistic.

1. Joseph L. Doob. Stochastic Processes. John Wiley and Sons, Inc. 1953
2. Feynman and Hibbs. Quantum Mechanics and Path Integrals. McGraw Hill. 1965

`