Statistical distance and Hilbert space*

W. K. Wootters
Center for Theoretical Physics, University of Texas, Austin, Texas 78712
(Received 2 September 1980)

A concept of "statistical distance" is defined between different preparations of the same quantum system, or in other words, between different rays in the same Hilbert space. Statistical distance is determined entirely by the size of statistical fluctuations occurring in measurements designed to distinguish one state from another. It is not related, *a priori*, to the usual distance (or angle) between rays. One finds, however, that these two kinds of distance are in fact the same, a result which depends on certain peculiarities of quantum mechanics.

I. INTRODUCTION: PHOTON POLARIZATION

If someone tosses a coin one hundred times and finds that "heads" occurs thirty times, he will conclude that the probability of heads is *roughly* 0.30 (the coin is weighted unusually). However, because of the unavoidable statistical fluctuations associated with a finite sample, he cannot know the value of this probability exactly. In the above example the probability of heads may well be around 0.26 or 0.34. The same thing happens in quantum measurements. If a finite ensemble of identically prepared quantum systems is analyzed by some fixed measuring device, the observed frequencies of occurrence of the various outcomes typically differ somewhat from the actual probabilities. Because of this statistical error, one cannot necessarily distinguish (in a fixed number of trials) between two slightly different preparations of the same quantum systems. We can say that two preparations are *indistinguishable* in a given number of trials if the difference in the actual probabilities is smaller than the size of a typical fluctuation.

In the present paper we use this idea of distinguishability to define a notion of distance, called "statistical distance," between quantum preparations. The definition involves counting the number of distinguishable states between two given states, when all states are analyzed by the same measuring device. Statistical distance is determined entirely by the size of statistical fluctuations, and has nothing particularly to do with the usual distance between pure states, i.e., the angle between rays in a Hilbert space. We shall find, however, that nature rather mysteriously makes these two kinds of distance identical. This will be our main result. It shows that there is a definite mathematical connection between the ubiquitous statistical fluctuations in the outcomes of measurements and the geometry of the set of states.

The concept of statistical distance is most easily introduced in terms of photons and polarizing filters. Imagine a beam of photons prepared by a polarizing filter and analyzed by a nicol prism. Let $\theta \in [0, \pi]$ be the angle by which the filter has been rotated (say, clockwise as viewed from the nicol prism) around the axis of the beam, starting from a standard position ($\theta = 0$) in which the filter's preferred axis is vertical. The filter is unmarked so that one cannot tell just by looking at it which axis is the preferred one.

Each photon, when it encounters the nicol prism, has exactly two options: to pass straight through the prism (call this the "yes" outcome) or to be deflected in a specific direction characteristic of the prism (the "no" outcome). Let us assume that the orientation of the nicol prism is fixed once and for all in such a way that vertically polarized photons always pass straight through. By counting how many photons yield each of the two possible outcomes, an experimenter can learn something about the value of θ via the formula $p = \cos^2\theta$, where p is the probability of yes.

Let us now suppose that the experimenter, in making his determination of the value of θ, has only a limited number of photons to work with, so that precisely n photons actually pass through the filter to be analyzed by the nicol prism. Then, because of the statistical fluctuations associated with a finite sample, the observed frequency of occurrence of yes is only an approximation to the actual probability of yes, the typical error being of the order of $n^{-1/2}$. More precisely, the experimenter's uncertainty (root-mean-square deviation) in the value of p is[1,2]

$$\Delta p = \left[\frac{p(1-p)}{n}\right]^{1/2}.$$

(This expression for Δp follows from elementary probability theory without any input from physics. The same formula would apply, for example, if one were trying to find the probability of heads of a weighted coin.) This uncertainty causes the experimenter to be uncertain of the value of θ by an amount

$$\Delta\theta = \left|\frac{dp}{d\theta}\right|^{-1} \Delta p = \left|\frac{dp}{d\theta}\right|^{-1} \left[\frac{p(1-p)}{n}\right]^{1/2} \quad (1)$$

Thus, we can associate with each value of θ a region of uncertainty, extending from $\theta - \Delta\theta$ to $\theta + \Delta\theta$, whose size could in principle depend on θ since both p and $dp/d\theta$ on the right-hand side of Eq. (1) depend on θ. Let us call two neighboring orientations θ and θ' *distinguishable in n trials* if their regions of uncertainty do not overlap, that is, if

$$|\theta - \theta'| \geq \Delta\theta + \Delta\theta' . \quad (2)$$

We now define the *statistical distance* $d(\theta_1, \theta_2)$ between any two orientations θ_1 and θ_2 to be

$$d(\theta_1, \theta_2) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \times [\text{maximum number of intermediate orientations each of which is distinguishable (in } n \text{ trials) from its neighbors}]. \quad (3)$$

In other words, the statistical distance is obtained essentially by counting the number of distinguishable orientations between θ_1 and θ_2. The factor $n^{-1/2}$ is included so that the limit will exist, the number of distinguishable orientations going as $n^{1/2}$. The statistical distance is intended to be a measure of how far apart θ_1 and θ_2 are in a statistical sense. It does not have anything to do, *a priori,* with the usual notion of distance (or angle) between θ_1 and θ_2, which is $|\theta_1 - \theta_2|$. We now show, however, that these two kinds of distance are in fact the same.

From Eqs. (1)–(3) we obtain the following expression for statistical distance in terms of the function $p(\theta)$ (assuming that $\theta_1 \leq \theta_2$):

$$d(\theta_1, \theta_2) = \frac{1}{\sqrt{n}} \int_{\theta_1}^{\theta_2} \frac{d\theta}{2\Delta\theta} = \int_{\theta_1}^{\theta_2} d\theta \frac{|dp/d\theta|}{2[p(1-p)]^{1/2}} .$$

Upon substituting the actual form of the probability law $p(\theta) = \cos^2\theta$ into this expression, we find that the statistical distance is

$$d(\theta_1, \theta_2) = \theta_2 - \theta_1 ;$$

that is, it is equal to the angle between the two orientations. This equality expresses the main result of this paper as it applies to the simple case of linear polarization of photons.

The fact that the proportionality constant between statistical distance and "actual distance" is unity is not particularly significant; it is due to our decision to divide by $\sqrt{n}$ in Eq. (3) rather than by some multiple of $\sqrt{n}$. However, the proportionality itself is nontrivial and depends on the fact that

$$\left|\frac{dp}{d\theta}\right| \propto [p(1-p)]^{1/2} , \quad (4)$$

something which would typically not be true if the probability law were different from $p(\theta) = \cos^2\theta$. In fact, the only periodic functions (with period 2π) satisfying Eq. (4) are those of the form

$$p(\theta) = \cos^2 \frac{m}{2}(\theta - \theta_0) ,$$

where m is an integer and θ_0 is a constant. Thus, if one were to demand of nature that the statistical distance be proportional to $|\theta_1 - \theta_2|$, the $\cos^2$ shape of the probability function would follow necessarily.

Another way of stating the above result is as follows. In the sequence of orientations given by $(\theta = 0, \theta = \epsilon, \theta = 2\epsilon, \ldots)$, all the orientations are equally distinguishable from their respective neighbors. This would follow trivially from rotational invariance if the nicol prism were allowed to be rotated. But we have assumed that the prism is fixed, and this is why the result is not empty.

In the above discussion, we defined statistical distance directly on the set of orientations of the polarizing filter. In the more general case—including measurements with more than two possible outcomes, also including elliptical polarizations in the case of photons—we must use a more roundabout approach. We first define in Sec. II the concept of statistical distance on probability space, a concept which applies to any probabilistic experiment, such as the throwing of dice. We then adapt this idea to quantum measurements in Sec. III. Finally, in Sec. IV we discuss the possible significance of the equivalence between statistical distance and angle in Hilbert space.

II. STATISTICAL DISTANCE ON PROBABILITY SPACE

The concept of statistical distance is quite independent of quantum mechanics and can be defined on any probability space. To emphasize this point, we now define the statistical distance between two differently weighted coins.

In a case such as this where there are exactly two possible outcomes, the probability space is one-dimensional, every coin being characterized by its probability of heads. The statistical distance $d(p_1, p_2)$ between two coins with probabilities p_1 and p_2 of heads is defined in a way analogous to that of the preceding section:

$$d(p_1, p_2) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \times [\text{maximum number of mutually distinguishable (in } n \text{ trials) intermediate probabilities}]. \tag{5}$$

Here two probabilities p and p' of heads are called distinguishable in n trials if

$$|p-p'| \geq \Delta p + \Delta p', \tag{6}$$

where, as before

$$\Delta p = \left[\frac{p(1-p)}{n}\right]^{1/2}. \tag{7}$$

From Eqs. (5)–(7) one obtains the following expression for the statistical distance:

$$d(p_1,p_2) = \int_{p_1}^{p_2} \frac{dp}{2[p(1-p)]^{1/2}} = \cos^{-1}(p_1^{1/2}p_2^{1/2} + q_1^{1/2}q_2^{1/2}),$$

where $q_1 = 1-p_1$ and $q_2 = 1-p_2$.

Thus, the statistical distance is not the same as the usual Euclidean distance on probability space, which would be $|p_1 - p_2|$. This is because probabilities near $\frac{1}{2}$ are more difficult to distinguish, owing to greater statistical fluctuations, than probabilities near 0 or 1 [cf. Eq. (7)]. Figure 1 shows a series of probabilities which are all equally spaced in the sense of statistical distance. The curves represent the distribution of the frequency of occurrence of yes for each of the special probabilities of yes. One obtains the statistical distance by counting the number of these curves that will "fit" between two given points.

The distance function defined by Eq. (5) was introduced, though not expressed in quite the same terms, in 1922 by Fisher in order to facilitate the analysis of "genetic drift."[3] Fisher's distance has since been used by other statisticians and geneticists, and has been called "genetic distance."[4-6]

We now wish to generalize the above definition to the case where there are more than two possible outcomes. This will allow us to speak of the statistical distance between two differently loaded dice, or, more important for our purpose, the statistical distance between two different preparations of a general quantum system.

Consider a probabilistic experiment (such as the throwing of a die) having exactly N possible outcomes. The probabilities $p_1, \ldots, p_N$ of the various outcomes can range over an $(N-1)$-dimensional probability space, restricted only by the requirements that the p_i's be non-negative and that they add to 1. (This space is most easily pictured in the case where $N=3$, for which the probability space is the triangular region shown in Fig. 2.) We will define the statistical distance between two points in this space in essentially the same way as before: It is proportional to the number of distinguishable intermediate points. But now "intermediate" does not have such a clear meaning, so we must introduce an extra step. We first define the statistical length of an arbitrary curve in probability space as follows: It is the maximum number of mutually distinguishable (in n trials) points *along the curve*, divided by $\sqrt{n}$, in the limit as $n\to\infty$. The statistical distance between two points is then the statistical length of the "shortest" such curve (in the sense of statistical length) connecting the two points. This idea is illustrated in Fig. 2.

To complete the definition, we need to say what it means for two points in this space to be dis-

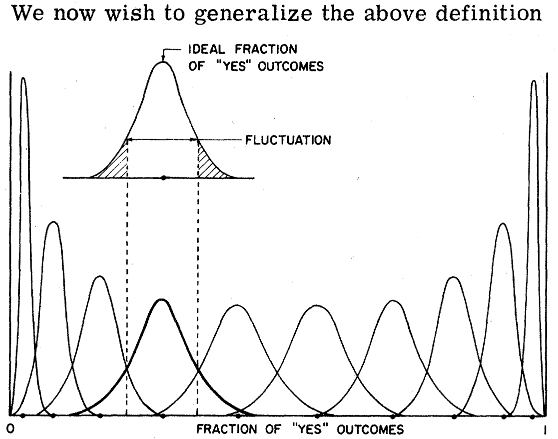

FIG. 1. These points are all equally spaced in the sense of statistical distance.

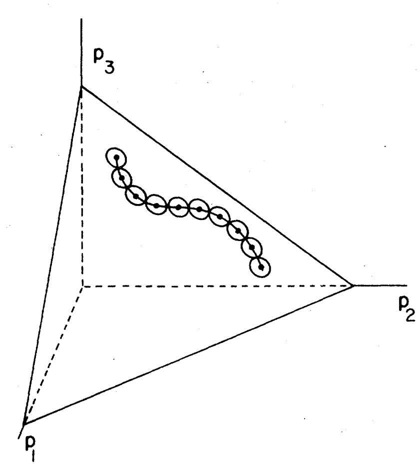

FIG. 2. Illustration of the definition of statistical length: proportional to the number of distinguishable points. Each point is shown with its region of uncertainty.

tinguishable in n trials. For a given set of probabilities $p_1, \ldots, p_N$, the actual frequencies of occurrence $\xi_1, \ldots, \xi_N$ of the outcomes (after n trials) are distributed according to a multinomial distribution, which can be approximated by a Gaussian when the number of trials is large[7]:

$$\rho(\xi_1, \ldots, \xi_N) \propto \exp\left[-\frac{n}{2}\sum_{i=1}^{N}\frac{(\xi_i - p_i)^2}{p_i}\right]. \qquad (8)$$

Let us define the *region of uncertainty* around the point $p = (p_1, \ldots, p_N)$ to be the set of all points $(\xi_1, \ldots, \xi_N)$ for which the exponent in Eq. (8) is less in absolute value than $\frac{1}{2}$. (This $\frac{1}{2}$ is chosen so that the present definition will agree with our earlier definition of distinguishability in the case where $N=2$.) Two points p and p' will be called *distinguishable in n trials* if their regions of uncertainty do not overlap. For large n, this will be the case if and only if

$$\frac{\sqrt{n}}{2}\left[\sum_{i=1}^{N}\frac{(\delta p_i)^2}{p_i}\right]^{1/2} > 1\,, \qquad (9)$$

where $\delta p_i = p_i - p_i'$. This completes the definition of statistical distance on the $(N-1)$-dimensional probability space.

To find an explicit expression for statistical distance, let $p^{(1)}$ and $p^{(2)}$ be two points in probability space, and let $p(t)$, $0 \leq t \leq 1$, parametrize a smooth curve lying in probability space and connecting these two points; thus $p(0) = p^{(1)}$ and $p(1) = p^{(2)}$. According to the above definition of statistical length and the criterion (9) for distinguishability, the statistical length of the curve $p(t)$ is

$$l = \frac{1}{2}\int_0^1 dt\left\{\sum_{i=1}^{N}\frac{1}{p_i(t)}\left[\frac{dp_i(t)}{dt}\right]^2\right\}^{1/2}.$$

One could now perform a variational calculation to find the shortest curve between $p^{(1)}$ and $p^{(2)}$, and thereby to find the statistical distance. But such a calculation is not necessary. If we change variables from p_i to x_i, defined by

$$x_i = p_i^{1/2}\,,$$

then the above expression for l becomes

$$l = \int_0^1 dt\left[\sum_{i=1}^{N}\left(\frac{dx_i}{dt}\right)^2\right]^{1/2}.$$

This is the usual Euclidean length of the curve in x space. The requirement that the curve $p(t)$ lie in probability space is expressed by the condition

$$1 = \sum_{i=1}^{N} p_i(t) = \sum_{i=1}^{N} x_i^2(t)\,.$$

Thus the curve $x(t)$ must lie on the unit sphere in x space.

The statistical distance between $p^{(1)}$ and $p^{(2)}$ is therefore the shortest distance along the unit sphere, between the points $x^{(1)}$ and $x^{(2)}$ defined by $x_i^{(1)} = (p_i^{(1)})^{1/2}$ and $x_i^{(2)} = (p_i^{(2)})^{1/2}$. This shortest distance is equal to the angle between the unit vectors $x^{(1)}$ and $x^{(2)}$ and is given by

$$d(p^{(1)}, p^{(2)}) = \cos^{-1}\left(\sum_{i=1}^{N} x_i^{(1)} x_i^{(2)}\right) = \cos^{-1}\left[\sum_{i=1}^{N}(p_i^{(1)})^{1/2}(p_i^{(2)})^{1/2}\right]. \qquad (10)$$

Equation (10) is our final expression for the statistical distance between $p^{(1)}$ and $p^{(2)}$. In a sense this is the most natural notion of distance on probability space, since it takes into account the actual difficulty of distinguishing different probabilistic experiments (e.g., differently weighted dice).

It is a straightforward matter to generalize the above definition of statistical distance to the case where the number of outcomes is countably infinite. Let the outcomes be labeled by $i = 1, \ldots, \infty$. As a first approximation, let us regard all the outcomes with $i = N+1, N+2, \ldots$, as just one outcome, and compute the statistical distance on that basis. If N is large enough, the probability $p_{N,\infty} = \sum_{i=N+1}^{\infty} p_i$ of this one outcome will be very small. In this approximation the statistical distance between two points $p^{(1)} = (p_1^{(1)}, p_2^{(1)}, \ldots)$ and $p^{(2)} = (p_1^{(2)}, p_2^{(2)}, \ldots)$ is, according to Eq. (10),

$$d_{\text{approx}}(p^{(1)}, p^{(2)}) = \cos^{-1}\left\{\sum_{i=1}^{N}\left[(p_i^{(1)})^{1/2}(p_i^{(2)})^{1/2}\right] + (p_{N,\infty}^{(1)})^{1/2}(p_{N,\infty}^{(2)})^{1/2}\right\}.$$

We define the exact statistical distance to be the limit of this quantity as N goes to infinity. Thus,

$$d(p^{(1)}, p^{(2)}) = \cos^{-1}\left[\sum_{i=1}^{\infty}(p_i^{(1)})^{1/2}(p_i^{(2)})^{1/2}\right]. \qquad (11)$$

Notice that the concept of a probability amplitude is quite foreign to the above derivation. The square roots of probability appearing in Eq. (11) come ultimately from Eq. (8), which in turn is based on the familiar combinatorial argument that leads to the multinomial distribution.

III. STATISTICAL DISTANCE EQUALS HILBERT-SPACE DISTANCE

We now wish to use the above ideas to define the statistical distance between two different preparations of a general quantum system. We will consider only preparations of pure states, which can be represented by rays in a Hilbert space. The question is therefore: What is the statistical distance between two rays $\psi^{(1)}$ and $\psi^{(2)}$?

We imagine the following experimental setup: There are two preparing devices, one of which prepares the state $\psi^{(1)}$ and the other of which prepares $\psi^{(2)}$. An experimenter who does not know which device prepares which state analyzes, by means of a fixed measuring device, the quantum systems (e.g., photons) emerging from the two preparing devices. The statistical distance between $\psi^{(1)}$ and $\psi^{(2)}$ is intended to be a measure of the number of distinguishable preparations "between" $\psi^{(1)}$ and $\psi^{(2)}$.

The new feature involved in treating quantum systems, as opposed to dice, is this: Whereas for dice there is only one possible experiment to perform (namely, rolling the die), for quantum systems there are many, one for each different analyzing device. Furthermore, two preparations may be more easily distinguished with one analyzing device than with another. For example, the vertical and horizontal polarizations of photons can be easily distinguished with an appropriately oriented nicol prism, but cannot be distinguished at all with a device whose eigenstates are the right- and left-handed circular polarizations. For this reason, we will speak of the statistical distance between two preparations $\psi^{(1)}$ and $\psi^{(2)}$ *with respect to* a particular measuring device. The *absolute* statistical distance between $\psi^{(1)}$ and $\psi^{(2)}$ is then defined as the largest such distance; that is, it is the statistical distance between $\psi^{(1)}$ and $\psi^{(2)}$ when they are analyzed by the most discriminating apparatus.

To translate this definition into mathematics, let $\phi_1, \ldots, \phi_N$ (N may be infinity) be the eigenstates of a measuring device A by which $\psi^{(1)}$ and $\psi^{(2)}$ are to be distinguished. Let us assume that these eigenstates are nondegenerate so that there are N distinct outcomes of each measurement. The probabilities of the various outcomes are $|(\phi_i, \psi^{(1)})|^2$ if the preparation is $\psi^{(1)}$ and $|(\phi_i, \psi^{(2)})|^2$ if the preparation is $\psi^{(2)}$. (Here the objects $\psi^{(i)}$ and ϕ_i should be interpreted as unit vectors belonging to the appropriate rays.) Thus, according to Eq. (11), the statistical distance between $\psi^{(1)}$ and $\psi^{(2)}$ with respect to the analyzing device A is

$$d_A(\psi^{(1)}, \psi^{(2)}) = \cos^{-1}\left[\sum_{i=1}^{N} |(\phi_i, \psi^{(1)})|\,|(\phi_i, \psi^{(2)})|\right]. \quad (12)$$

One can easily convince oneself that this quantity achieves its maximum value if one of the eigenstates of A (say, ϕ_1) is the same as $\psi^{(1)}$ (or $\psi^{(2)}$). In that case only the $i=1$ term contributes to the sum in Eq. (12), and we find that the absolute statistical distance is

$$d(\psi^{(1)}, \psi^{(2)}) = \cos^{-1}|(\psi^{(1)}, \psi^{(2)})|.$$

This is our main result: that the (absolute) statistical distance between two preparations is equal to the angle in Hilbert space between the corresponding rays.

The angle in Hilbert space is the only Riemannian metric on the set of rays, up to a constant factor, which is invariant under all unitary transformations, that is, under all possible time evolutions. In this sense it is a natural metric on the set of states. It is interesting that the same metric arises from quite another starting point, namely, the analysis of statistical fluctuations in a finite sequence of measurements.

IV. DISCUSSION

The above result is in a way very appealing. It is as if nature defines distance between states by counting the number of distinguishable intermediate states. One can hardly imagine a more natural way of defining distance, and yet we cannot claim at this point to understand physically this connection between statistics and geometry.

We saw in the case of photon polarization that the $\cos^2$ shape of the probability law follows from the requirement that the statistical distance be proportional to the usual distance (or angle) between orientations of the filter. It is interesting to ask whether this kind of deduction could be made more generally. By requiring that the distance between states of a general quantum system be determined by the number of distinguishable intermediate states, could we conclude that the set of states as a whole must have the geometric structure of the set of rays in a complex vector space? We have seen that the concept of statistical distance *is* capable of converting a flat probability space into a section of the unit sphere in a *real* N-dimensional vector space, with probabilities being the squares of real amplitudes. However, there is nothing in the above analysis which would tell us, if we did not already know, that the actual set of states exists in a complex space. Therefore, if statistical distance is involved in determining the geometry of this set, it is not the whole story.

Nevertheless, the equivalence between statistical distance and Hilbert-space distance remains surprising and raises the interesting possibility that statistical fluctuations in the outcomes of measurements might be partly responsible for the Hilbert-space structure of quantum mechanics, a structure which seems too rich not to have a deeper foundation.[8] These statistical fluctuations are as basic as the fact that quantum measurements are probabilistic; so it is not inconceivable that they could play such a fundamental role.

ACKNOWLEDGMENTS

I would like to thank Professor J. A. Wheeler and Professor L. E. Reichl for many valuable discussions. I would also like to thank the Center for Statistical Mechanics of the University of Texas at Austin and the National Science Foundation (Grant No. PHY 7826592) for supporting this work.

---

*Based on the author's doctoral dissertation, University of Texas at Austin, May 1980.

[1] M. Fisz, *Probability Theory and Mathematical Statistics*, 3rd ed. (Wiley, New York, 1963), Sec. 5.2.

[2] Strictly speaking, one cannot define a root-mean-square deviation without choosing a specific *a priori* distribution of the probability of "yes." However, for sufficiently large n, our expression for Δp is valid regardless of the choice of *a priori* distribution, as long as this distribution is smooth and nonvanishing in the interval [0,1]. The problem of *a priori* distributions is discussed, for example, by S. Sykora, J. Stat. Phys. 11, 17 (1974).

[3] R. A. Fisher, Proc. R. Soc. Edinburgh 42, 321 (1922).

[4] F. Mosteller and J. W. Tukey, J. Am. Stat. Assoc. 44, 174 (1949).

[5] L. L. Cavalli-Sforza and F. Conterio, Atti Assoc. Genet. Ital. 5, 333 (1960).

[6] M. Kimura, *Diffusion Models in Population Genetics* (Methuen, London, 1962), pp. 23–25. In this and the preceding reference the genetic distance is defined on spaces of more than one dimension and is the same as in our Eq. (10).

[7] B. V. Gnedenko, *The Theory of Probability* (Chelsea, New York, 1962), p. 85.

[8] The view that such a deeper foundation is within our grasp has been expressed by J. A. Wheeler, *Frontiers of Time* (North-Holland, Amsterdam, 1979).