International Journal of Theoretical Physics, Vol. 26, No. 3, 1987

Probability Space of Wave Functions

Silviu Guiasu[1]

Received November 27, 1986

It is well known that in quantum mechanics, when the mean value of an observable is given, entropy maximization (von Neumann, Born, Jaynes) can be successfully applied for constructing a probability distribution on the set of possible values of that observable. In this paper, the entropy maximization technique is extended to the complex domain in order to construct an unbiased probability measure on the set of all wave functions. In particular, a justification and a generalization of the Wiener-Siegel probability distribution of Gaussian type in the differential space of wave functions are given.

1. INTRODUCTION

In classical statistical mechanics the possible states are points in a Euclidean space, the phase space, and when the exact state is not known, a probability measure is constructed on the Borel-measurable set of states. In quantum mechanics, the state of the system is a wave function with complex coefficients, which, according to Born, will in general permit statements as to the probability of finding different values of some quantities of interest. We may encounter situations where we do not know the precise state of the system, namely its wave function at time t_0. In such a case, as in classical statistical mechanics, we have to organize the state space, i.e., the space of wave functions, as a probability space and make probabilistic predictions on the possible wave functions, which themselves allow only probabilistic interpretations. There is here a double intervention of probability. As noticed by Messiah (1969), the so-called density matrix introduced by von Neumann (1932) in order to solve this problem plays a role only somewhat similar to that of the probability density function from classical statistical mechanics. On the other hand, in a series of papers Wiener and Siegel (1953, 1955, 1966), inspired by the Brownian motion stochastic process, *postulated* a probability distribution of Gaussian type on the set

[1]Department of Mathematics, York University, Toronto, Ontario M3J 1P3, Canada.

0020-7748/87/0300-0275$05.00/0

of wave functions, in an attempt to build up a quantum statistical mechanics closer to the approach used in classical statistical mechanics for organizing the whole state space as a probability space.

The aim of this paper is to construct an unbiased probability measure on the space of wave functions subject to given data, available at the macroscopic scale, by extending the use of entropy maximization to the complex domain. As is well known, the entropy maximization technique allows us, in general, to construct the most random or, better, the most unbiased probability distribution compatible with given mean values of random variables. After the major contributions of von Neumann (1932), Born (1969), and Jaynes (1957) in this area, many applications of this technique have been made in recent years, as can be seen, for instance, in Levine and Tribus (1979) or Smith and Grandy (1985). Generally, these contributions have focused mainly on how to construct a probability distribution on the set of possible values of one or several *observables* of a quantum system rather than on the set of wave functions that are the possible states of the quantum system.

Let $\{x_k\}$ be the complex components of an arbitrary wave function ψ with respect to a complete orthonormal set of eigenfunctions $\{\phi_k\}$ corresponding to the eigenvalues $\{u_k\}$ of an observable U. We put $x_k = x_{k,1} + ix_{k,2} = r_k \exp(i\theta_k)$. By performing measurements at the macroscopic scale we can generally get only mean values of random variables. In Section 2, supposing that we know the mean probability $|c_k|^2$ that the observable U takes on the value u_k, the principle of maximum entropy is successively applied for constructing the most unbiased probability distribution on the possible values of $|x_k|^2$, r_k, θ_k, and of the pairs (r_k, θ_k) and $(x_{k,1}, x_{k,2})$. These results are used in Section 3 for inducing a probability measure on the space of complex sequences $\{x_k\}$. Section 4 deals with the common case when the only available information is supplied by the mean value of the observable U. An example involving the one-dimensional harmonic oscillator is discussed. Once the space of wave functions is organized as a probability space, in Section 5, the mean and the variance of the mean value of an observable are obtained. In Section 6, bounds for the random norm of ψ are given. In Section 7, connections are made between the present approach and the statistical ensembles from Wiener and Siegel's differential space. The polychotomic algorithm of Wiener and Siegel is obtained in a more general context. The last section contains conclusions.

2. SOME USEFUL PROBABILITY DISTRIBUTIONS

Before applying entropic optimization techniques in quantum statistical mechanics let us make some preliminary remarks. If $\{\phi_k, k = 0, 1, \ldots\}$ is a

complete orthonormal set of eigenfunctions corresponding to the eigenvalues $\{u_k, k=0, 1, \ldots\}$ of the observable U and if the normalized wave function ψ is written as

$$\psi = \sum_k c_k \phi_k \tag{1}$$

then the expected value of U corresponding to the state ψ has the form

$$\langle U \rangle \psi = \sum_k u_k |c_k|^2$$

which entitles us to interpret $|c_k|^2$ as being the probability that the observable U takes on the value u_k. Suppose that at the time of preparation t_0 the dynamical state of the system is known incompletely and we assign to it a statistical mixture of a finite or countable set of wave functions $\{\psi_j, j = 0, 1, \ldots\}$, not necessarily normalized, with the corresponding statistical weights $\{p_j, j=0, 1, \ldots\}$ such that

$$p_j > 0 \ (j=0, 1, \ldots), \qquad \sum_j p_j = 1$$

Taking the eigenfunctions $\{\phi_k, k=0, 1, \ldots\}$ of the observable U as basis, we have

$$\psi_j = \sum_k c_{j,k} \phi_k$$

and, if $\langle \cdot | \cdot \rangle$ denotes the inner product, let us define

$$|c_k|^2 = \sum_j p_j \frac{|c_{j,k}|^2}{\langle \psi_j | \psi_j \rangle} \tag{2}$$

In particular, the set of wave functions $\{\psi_j, j=0, 1, \ldots\}$ may be just $\{\phi_k, k=0, 1, \ldots\}$, in which case $c_{j,j}=1$ and $c_{j,k}=0$ if j and k are different, which implies

$$|c_k|^2 = p_k$$

Let us take an arbitrary wave function

$$\psi = \sum_k x_k \phi_k = \sum_k r_k \exp(i\theta_k) \ \phi_k \tag{3}$$

where

$$r_k = |x_k| \qquad \text{and} \qquad \theta_k = \arg x_k$$

Of course, the numbers $\{|c_k|^2\}$ do not determine the state of the system at time t_0. They can, however, determine a probability distribution on the state space of the system. By performing measurements at the macroscopic scale wa can generally get only mean values of random variables. Let us

look at the number $|c_k|^2$ as being the average (or expected) value of $|x_k|^2$ or r_k^2. Using the principle of maximum entropy (PME) we can construct the most random (the most unbiased, the most uniform) probability distribution on the possible values of $|x_k|^2$ subject to the expected value $|c_k|^2$. After that we obtain successively the probability distribution on the values taken on by the module r_k, the argument θ_k, the pair (r_k, θ_k), and the pair $(x_{k,1}, x_{k,2})$ of the real and imaginary components of x_k $(x_k = x_{k,1} + ix_{k,2})$. From here on we apply the standard technique for constructing the product measure on the set of sequences $\{(x_{1,1}, x_{1,2}), (x_{2,1}, x_{2,2}), \dots\}$, each such sequence corresponding to a well-defined wave function (state) of the system,

$$\psi = \sum_k (x_{k,1} + ix_{k,2})\phi_k$$

The following propositions will realize, step by step, this program.

Proposition 1. The maximum entropy probability distribution on the possible values of the square of the module r_k^2 compatible with the expected value $|c_k|^2$ has the exponential probability density

$$f_k(y) = \begin{cases} \exp(-y/|c_k|^2)/|c_k|^2 & \text{if} \quad y > 0 \\ 0 & \text{elsewhere} \end{cases} \tag{4}$$

The mean of this probability distribution is $|c_k|^2$ and the variance $|c_k|^4$.

Proof. The computation is quite standard. We want to solve the convex nonlinear program with two equality constraints:

$$\max_{f_k} \left[-\int_0^{+\infty} f_k(y) \ln f_k(y)\, dy \right] \tag{5}$$

subject to

$$\int_0^{+\infty} f_k(y)\, dy = 1 \tag{6}$$

$$\int_0^{+\infty} y f_k(y)\, dy = |c_k|^2 \tag{7}$$

Let $H(f_k)$ be the entropy of the probability density f_k whose expression is given in (5) and α and β be Lagrange multipliers corresponding to the constraints (6) and (7). Taylor's formula tells us that for any $t > 0$ there is τ, depending on t, between 1 and t, such that

$$G(t) = t \ln t = (t-1) + (t-1)^2/(2\tau)$$

Using this formula, we get

$$\begin{aligned} &-H(f_k)+\alpha\cdot 1+\beta|c_k|^2 \\ &=\int_0^{+\infty}\exp(-\alpha-\beta y)\; G[f_k(y)\exp(\alpha+\beta y)]\,dy \\ &\geq\int_0^{+\infty}\exp(-\alpha-\beta y)\,[f_k(y)\exp(\alpha+\beta y)-1]\,dy \\ &=1-\int_0^{+\infty}\exp(-\alpha-\beta y)\,dy \end{aligned} \tag{8}$$

with equality if and only if

$$f_k(y)=\exp(-\alpha-\beta y), \qquad y>0 \tag{9}$$

in which case the lower bound in (8) is equal to zero. Introducing (9) into (6) and (7), after an elementary integration we obtain (4).

Proposition 2. The values taken on by the module r_k are distributed according to the Weibull probability density

$$g_k(z)=\begin{cases} 2z\exp(-z^2/|c_k|^2)/|c_k|^2 & \text{if} \quad z>0 \\ 0 & \text{elsewhere} \end{cases} \tag{10}$$

Its mean is $\pi^{1/2}|c_k|/2$ and the variance $(1-\pi/4)|c_k|^2$.

Proof. As r_k^2 is distributed according to (4), we obtain for r_k the probability density

$$g_k(z)=f_k(z^2)\cdot 2z=2z\exp(-z^2/|c_k|^2)/|c_k|^2, \qquad z>0$$

which is the Weibull distribution

$$g(z)=\begin{cases} \alpha\beta z^{\beta-1}\exp(-\alpha z^\beta), & \text{if} \quad z>0 \\ 0 & \text{elsewhere} \end{cases}$$

with the parameters $\alpha=1/|c_k|^2$ and $\beta=2$. Its mean is

$$\mu=\alpha^{-1/\beta}\Gamma(1+1/\beta)=\pi^{1/2}|c_k|/2\simeq 0.8862|c_k|$$

and the variance

$$\sigma^2=\alpha^{-2/\beta}[\Gamma(1+2/\beta)-\Gamma^2(1+1/\beta)]=(1-\pi/4)|c_k|^2\simeq 0.2146|c_k|^2$$

Proposition 3. The maximum entropy probability distribution on the possible values of the argument θ_k has the uniform probability density

$$h_k(s)=\begin{cases} 1/(2\pi) & 0\leq s\leq 2\pi \\ 0 & \text{elsewhere} \end{cases} \tag{11}$$

Its mean is π and the variance is equal to $\pi^2/3$.

Proof. We can follow the same technique as in the proof of Proposition 1 in order to maximize $H(h_k)$ subject to the only constraint

$$\int_0^{2\pi} h_k(s)\, ds = 1$$

and we obtain (11). The mean and the variance may be calculated in the usual way.

Proposition 4. Since the module r_k and the argument θ_k are independent, their joint probability density is

$$\xi_k(z, s) = g_k(z)h_k(s) = \begin{cases} z \exp(-z^2/|c_k|^2)/\pi|c_k|^2 & \text{for} \quad z>0,\ 0\le s\le 2\pi \\ 0 & \text{elsewhere} \end{cases} \tag{12}$$

Proposition 5. Each component of the independent pair $(x_{k,1}, x_{k,2})$, where $x_k = x_{k,1} + ix_{k,2} = r_k \exp(i\theta_k)$, is normally distributed with the mean zero and the variance $|c_k|^2/2$, its probability density being

$$\eta_k(v) = \exp(-v^2/|c_k|^2)/\pi^{1/2}|c_k| \tag{13}$$

Proof. Let us denote by ζ_k the joint probability density of the pair $(x_{k,1}, x_{k,2})$, where

$$x_k = x_{k,1} + ix_{k,2} = r_k \exp(i\theta_k)$$

Since

$$x_{k,1} = r_k \cos\theta_k, \qquad x_{k,2} = r_k \sin\theta_k$$

taking (12) into account, we have

$$\begin{aligned} \zeta_k(x_{k,1}, x_{k,2}) &= \xi_k(r_k, \theta_k)\left|(\partial(x_{k,1}, x_{k,2})/\partial(r_k, \theta_k))^{-1}\right| \\ &= \xi_k(r_k, \theta_k)/r_k \\ &= \exp(-r_k^2/|c_k|^2)/\pi|c_k|^2 \\ &= \exp[-(x_{k,1}^2 + x_{k,2}^2)/|c_k|^2/\pi|c_k|^2 \\ &= [\exp(-x_{k,1}^2/|c_k|^2)/\pi^{1/2}|c_k|][\exp(-x_{k,2}^2)/|c_k|^2)/\pi^{1/2}|c_k|] \end{aligned} \tag{14}$$

which is the product of two normal marginal probability distributions $N(0, |c_k|^2/2)$.

Remark. When $|c_k|^2 = 1$ we obtain

$$f_k(y) = \exp(-y) \quad (y>0); \qquad g_k(z) = 2z\exp(-z^2) \quad (z>0)$$

$$\eta_k(v) = \exp(-v^2)/\pi^{1/2}$$

the joint probability density ζ_k of the pair $(x_{k,1}, x_{k,2})$ being the product of two normal distributions $N(0, \frac{1}{2})$. In such a case, the joint probability density

$$\xi_k(z, s) = \begin{cases} z \exp(-z^2)/\pi & \text{if} \quad z > 0, \quad 0 \leq s \leq 2\pi \\ 0 & \text{elsewhere} \end{cases}$$

for (r_k, θ_k), or the joint probability density

$$\zeta_k(v_1, v_2) = \exp[-(v_1^2 + v_2^2)]/\pi$$

for $(x_{k,1}, x_{k,2})$ characterizes the probability distribution of the possible states of the system having the form

$$\psi = x_k \phi_k$$

where

$$x_k = x_{k,1} + i x_{k,2} = r_k \exp(i\theta_k)$$

3. A PROBABILITY MEASURE ON THE STATE SPACE

Let us introduce the countable set

$$\mathbf{A} = \{(k, 1), (k, 2); k = 0, 1, 2, \ldots\}$$

and the collection $\mathbf{R}^{\mathbf{A}}$ of all real-valued functions $\omega(\alpha)$, $\alpha \in \mathbf{A}$, defined on $\mathbf{A}$. Let $p_{\alpha_1,\ldots,\alpha_n}$ be the projection of $\mathbf{R}^{\mathbf{A}}$ onto $\mathbf{R}^n$ corresponding to a finite sequence $\{\alpha_1, \ldots, \alpha_n\}$ of distinct elements of $\mathbf{A}$ defined by

$$p_{\alpha_1,\ldots,\alpha_n}(\omega) = [\omega(\alpha_1), \ldots, \omega(\alpha_n)] \in \mathbf{R}^n$$

and let us introduce the σ-field of all Borel cylinders in $\mathbf{R}^{\mathbf{A}}$ with index $\{\alpha_1, \ldots, \alpha_n\}$,

$$I_{\alpha_1,\ldots,\alpha_n} = \{p^{-1}_{\alpha_1,\ldots,\alpha_n}(B), B \in \mathbf{B}^n\}$$

where $\mathbf{B}^n$ is the σ-field of Borel sets in $\mathbf{R}^n$. On the field

$$I = \cup\, I_{\alpha_1,\ldots,\alpha_n}$$

where the union is taken over all finite sequences of distinct elements of $\mathbf{A}$, we define a set function η as follows. If $E \in I$ and $E \in I_{\alpha_1,\ldots,\alpha_n}$, having the form

$$E = p_{\alpha_1,\ldots,\alpha_n}(B), \qquad B \in \mathbf{B}^n$$

we put

$$\eta(E) = \int_{\mathbf{R}^n} \chi_B(x) \eta_{\alpha_1,\ldots,\alpha_n}(x)\, dx$$

where χ_B is the indicator of the set B and for $\alpha_1=(k_1, l_1), \ldots, \alpha_n=(k_n, l_n)$ and $v=(v_1, \ldots, v_n)$, we have, for any $l_j \in \{0, 1\}$,

$$\eta_{\alpha_1,\ldots,\alpha_n}(v)=\prod_{j=1}^{n}\eta_{\alpha_j}(v_j)=\pi^{-n/2}|c_{k_1}|^{-1}\cdots|c_{k_n}|^{-1}\exp\left(-\sum_{j=1}^{n}(v_j^2/|c_{k_j}|^2)\right) \quad (15)$$

The set function η is well defined on I and, according to the Kolmogorov extension theorem, η can be extended uniquely to be a probability measure on the σ-field $\sigma(I)$ generated by I.

On the probability space $\{\mathbf{R}^{\mathbf{A}}, \sigma(I), \eta\}$ we can define a system of random variables

$$\mathbf{X}=\{X(\alpha, \cdot), \alpha \in \mathbf{A}\}$$

by

$$X(\alpha, \omega)=\omega(\alpha) \qquad \text{for} \quad \omega \in \mathbf{R}^{\mathbf{A}}$$

Then, for $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbf{A}$ and $B \in \mathbf{B}^n$ we have

$$\begin{aligned}&\eta(\{\omega;\ \omega \in \mathbf{R}^{\mathbf{A}}, [X(\alpha_1, \omega), \ldots, X(\alpha_n, \omega)] \in B\})\\ &=\eta(\{\omega;\ \omega \in \mathbf{R}^{\mathbf{A}}, [\omega(\alpha_1), \ldots, \omega(\alpha_n)] \in B\})\\ &=\eta(p^{-1}_{\alpha_1,\ldots,\alpha_n}(B))=\int_{\mathbf{R}^n}\chi_B(v)\eta_{\alpha_1,\ldots,\alpha_n}(v)\,dv\end{aligned}$$

so that $\mathbf{X}$ is a Gaussian system of independent random variables.

The set $\mathbf{R}^{\mathbf{A}}$ is larger than the set of wave functions. An element

$$\omega=(x_{0,1}, x_{0,2}, x_{1,1}, x_{1,2}, \ldots, x_{k,1}, x_{k,2}, \ldots) \in \mathbf{R}^{\mathbf{A}} \quad (16)$$

corresponds to the state

$$\psi=\sum_k (x_{k,1}+ix_{k,2})\phi_k$$

if and only if

$$\sum_k |x_k|^2=\sum_k r_k^2<+\infty \quad (17)$$

where

$$x_k=x_{k,1}+ix_{k,2}=r_k\exp(i\theta_k)$$

We prefer, however, to look at any sequence (16) as representing a possible (extended) state of the system. Writing, for brevity,

$$\psi=(x_{0,1}, x_{0,2}, x_{1,1}, x_{1,2}, \ldots, x_{k,1}, x_{k,2}, \ldots) \in \mathbf{R}^{\mathbf{A}}$$

we denote the integral of a function $f(\psi)$ with respect to the measure η, namely

$$\int_{\mathbf{R}^{\mathbf{A}}} f(\psi)\, d\eta(\psi)$$

by $\mathbf{E}(f(\psi))$.

Of course, not always does

$$\sum_k |x_k|^2 = 1$$

but according to the next proposition, this equality is true in the mean.

Proposition 6.

$$\mathbf{E}(\langle\psi|\psi\rangle) = 1 \tag{18}$$

Proof. Using (13), we get

$$\begin{aligned}\mathbf{E}(\langle\psi|\psi\rangle) &= \int_{\mathbf{R}^{\mathbf{A}}} \langle\psi|\psi\rangle\, d\zeta(\psi)\\ &= \sum_k \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x_{k,1}^2 + x_{k,2}^2)\\ &\quad \times \exp[-(x_{k,1}^2 + x_{k,2}^2)/|c_k|^2]/\pi|c_k|^2\, dx_{k,1}\, dx_{k,2}\\ &= \sum_k |c_k|^2 = 1\end{aligned}$$

Remark. If we denote

$$\Theta(z) = \int_0^z \exp(-x^2/2)/(2\pi)^{1/2}\, dz$$

then the probability of having a state belonging to the set

$$\left\{\psi;\ \psi = \sum_k (x_{k,1} + ix_{k,2})\phi_k,\ a_{k,j} \le x_{k,j} \le b_{k,j}\ (j = 1, 2), \text{ for any } k\right\}$$

is equal to

$$\prod_k \prod_j [\Theta(2^{1/2} b_{k,j}/|c_k|) - \Theta(2^{1/2} a_{k,j}/|c_k|)]$$

Following the steps mentioned above, we organize the state space as a probability space at time t_0. On the other hand, the Schrödinger equation induces a strictly deterministic flow $S(t_0, t)$ on the state space $\mathbf{R}^{\mathbf{A}}$. Supposing that the transformation $S(t_0, t)$ is $\sigma(I)$-measurable, the initial probability space $\{\mathbf{R}^{\mathbf{A}}, \sigma(I), \eta\}$ at time t_0, is transformed into the new probability space $\{\mathbf{R}^{\mathbf{A}}, \sigma(I), \eta S^{-1}(t_0, t)\}$ at time t $(t > t_0)$.

4. THE CASE WHEN THE MEAN VALUE OF AN OBSERVABLE IS GIVEN

The above considerations are based on the fact that initially we know, from observations, the values of $|c_k|^2$ $(k = 0, 1, \ldots)$. These values have been interpreted as the mean values of the probabilities that the observable U takes on the eigenvalues u_k $(k = 0, 1, \ldots)$ corresponding to the eigenfunctions ϕ_k $(k = 0, 1, \ldots)$. Such an interpretation is in accordance both with statistical mechanics and with statistical inference, which have taught us that the only way of coping with random events and fluctuations is to start from mean values estimated at the macroscopic level. Sometimes the numbers $|c_k|^2$ cannot be directly estimated and the only available information is given by the mean value of the observable. Suppose that at time t_0 we know the mean value $\langle U \rangle$ of the observable U. Practically, we measure the value taken on by U for a set of indentical systems and we use these observations in order to estimate the mean value $\langle U \rangle$ of U, getting generally a confidence interval for $\langle U \rangle$, or testing a statistical hypothesis about $\langle U \rangle$. (This is the simplest strategy to follow; in order to estimate the mean probabilities $|c_k|^2$, $k = 0, 1, \ldots$, using the relative frequencies of the values u_k, $k = 0, 1, \ldots$, taken on by the observable U for a set of identical systems, generally we need much more observations than for estimating the mean value $\langle U \rangle$ of U.) In such a case, since

$$\langle U \rangle = \sum_k u_k |c_k|^2 \tag{19}$$

there are many mean probability distributions $\{|c_k|^2, k = 0, 1, \ldots\}$ compatible with the given mean value $\langle U \rangle$. In order to make a choice, a natural strategy would be to select from this set of feasible mean probability distributions the most random one (the most unbiased one), treating the possible values of U as uniformly as possible subject to the constraint (19). Therefore, we can apply the standard form of the principle of maximum entropy, used by von Neumann (1932), Born (1969), and Jaynes (1957), in order to determine the mean probability distribution $\{|c_k|^2, k = 0, 1, \ldots\}$ that maximizes the entropy

$$H(|c|^2) = -\sum_k |c_k|^2 \ln |c_k|^2$$

subject to

$$\sum_k |c_k|^2 = 1$$

$$\sum_k u_k |c_k|^2 = \langle U \rangle$$

Using the same steps as in the proof of Proposition 1, we obtain the known solution (Gibbs's discrete canonical distribution)

$$|c_k|^2 = \exp(-\beta u_k)/\Phi(\beta), \qquad k = 0, 1, \ldots \tag{20}$$

where

$$\Phi(\beta) = \sum_k \exp(-\beta u_k)$$

with β satisfying the equation

$$\Phi'(\beta)/\Phi(\beta) = -\langle U \rangle \tag{21}$$

In the particular, but important, case when

$$u_k = u_0 + ku, \qquad k = 0, 1, \ldots$$

equation (21) can be easily solved and we get

$$|c_k|^2 = u(\langle U \rangle - u_0)^k/(\langle U \rangle + u - u_0)^{k+1}, \qquad k = 0, 1, \ldots$$

and this expression has to be introduced into the formulas (4)-(15).

For a one-dimensional harmonic oscillator, for instance, if U is the energy of the oscillator, we have

$$u_0 = \tfrac{1}{2}h\nu, \qquad u = h\nu$$

where ν is the frequency of the oscillator and h is the Planck constant. In such a case

$$|c_k|^2 = h\nu(\langle U \rangle - h\nu/2)^k/(\langle U \rangle + h\nu/2)^{k+1}, \qquad k = 0, 1, \ldots$$

Replacing these values in (12) and (14), we get the maximum entropy probability distribution of the pairs $\{(r_k, \theta_k), k = 0, 1, \ldots\}$ and the corresponding probability distribution of the pairs $\{(x_{k,1}, x_{k,2}), (k = 0, 1, \ldots)\}$ of the coefficients

$$x_k = r_k \exp\{i\theta_k\} = x_{k,1} + ix_{k,2}$$

of the possible states

$$\psi = \sum_k x_k \phi_k$$

where $\{\phi_k, k = 0, 1, \ldots\}$ are the eigenfunctions of the energy of the harmonic oscillator

$$\phi_k(t) = N_k \exp(-\alpha t^2/2) H_k(\alpha^{1/2} t)$$

where

$$\alpha = 4\pi^2 \nu m/h, \qquad N_k = [(\alpha/\pi)^{1/2}/(2^k k!)]^{1/2}$$

m is the mass of the oscillator, and H_k, denotes the Hermite polynomials.

5. THE MEAN OF THE MEAN VALUE OF AN OBSERVABLE

Let V be a Hermitic operator corresponding to an observable denoted by the same letter. The mean value of V corresponding to the state ψ is

$$\langle V\rangle\psi = \langle\psi|\, V\psi\rangle / \langle\psi|\,\psi\rangle$$

In our approach, since ψ is the generic element of a probability space, the mean $\langle V\rangle$ becomes a random variable. For any subset $B \in I$ of states we define the mean value of V corresponding to the subset B, or on B, as being

$$\langle V\rangle(B) = \mathbf{E}(\langle\psi|\, V\psi\rangle\chi_B(\psi)) / \mathbf{E}(\langle\psi|\,\psi\rangle\chi_B(\psi))$$

if the two integrals exist. In particular, if $B = \mathbf{R}^{\mathbf{A}}$, then we obtain the following result.

Proposition 7. The mean value of V on $\mathbf{R}^{\mathbf{A}}$ (the mean of the mean) is

$$\langle V\rangle = \langle V\rangle(\mathbf{R}^{\mathbf{A}}) = \sum_k \langle\phi_k|\, V\phi_k\rangle |c_k|^2 \tag{22}$$

Proof. Taking (18) into account, we get

$$\begin{aligned}
\langle V\rangle = \langle V\rangle(\mathbf{R}^{\mathbf{A}}) &= \mathbf{E}(\langle\psi|\, V\psi\rangle) = \int_{\mathbf{R}^{\mathbf{A}}} \langle\psi|\, V\psi\rangle\, d\eta(\psi) \\
&= \int_0^{+\infty} \int_0^{+\infty} \sum_{k,l} (x_{k,1} - ix_{k,2})(x_{l,1} + ix_{l,2}) \langle\phi_k|\, V\phi_l\rangle \\
&\quad \times \prod_{s=0}^{n} \exp[-(x_{s,1}^2 + x_{s,2}^2)/|c_s|^2]/\pi|c_s|^2\, dx_{s,1}\, dx_{s,2} \\
&= \sum_k \langle\phi_k|\, V\phi_k\rangle \sum_{j=1}^{2} \int_{-\infty}^{+\infty} x_{k,j}^2 \exp(-x_{k,j}^2/|c_k|^2)/\pi^{1/2}|c_k|\, dx_{k,j} \\
&= \sum_k \langle\phi_k|\, V\phi_k\rangle |c_k|^2
\end{aligned}$$

Remark. If $\{\phi_k, k = 0, 1, \ldots\}$ are eigenfunctions of V and if

$$V\phi_k = v_k\phi_k, \qquad k = 0, 1, \ldots \tag{23}$$

then

$$\langle V\rangle = \sum_k v_k |c_k|^2 \tag{24}$$

Proposition 8. The variance of V is

$$\sigma^2 = \langle V^2\rangle - (\langle V\rangle)^2 = \sum_k \langle V\phi_k|\, V\phi_k\rangle |c_k|^2 - \left(\sum_k \langle\phi_k|\, V\phi_k\rangle |c_k|^2\right)^2 \tag{25}$$

Proof. Indeed,

$$\langle V^2\rangle = \mathbf{E}(\langle\psi|\, V^2\psi\rangle) = \mathbf{E}(\langle V\psi|\, V\psi\rangle) = \int_{\mathbf{R}^A} \langle V\psi|\, V\psi\rangle\, d\eta(\psi)$$

$$= \sum_k \langle V\phi_k|\, V\phi_k\rangle \int_{-\infty}^{+\infty} (x_{k,1}^2 + x_{k,2}^2)$$

$$\exp[-(x_{k,1}^2 + x_{k,2}^2)/|c_k|^2]/\pi|c_k|^2\, dx_{k,1}\, dx_{k,2}$$

$$= \sum_k \langle V\phi_k|\, V\phi_k\rangle |c_k|^2$$

Taking into account (22), we get (25).

Remark. When the equalities (23) hold, then

$$\sigma_V^2 = \sum_k v_k^2 |c_k|^2 - \left(\sum_k v_k |c_k|^2\right)^2$$

Generally, we can assign only average properties or expected values to a system. The above statistical approach on the space of possible states offers a more elastic prediction of the mean value of an observable V. Indeed, from the last two propositions, Chebychev's inequality shows that the probability of having

$$|\langle V\rangle\psi - \sum_k \langle\phi_k|\, V\phi_k\rangle |c_k|^2| < t\sigma_V^2$$

is larger than $1 - t^{-2}$, where σ_V^2 is given by (25).

6. STATES REPRESENTED BY FINITE SUMS

Sometimes in practice we are dealing only with a finite number of eigenfunctions, i.e., we consider states of the form

$$\psi = \sum_{k=0}^{n} x_k \phi_k$$

In such a case we can compute the probability of some events of interest. Let us denote by X_k the random variable representing the square of the module of the coefficient x_k of ϕ_k. As we know, X_k is exponentially distributed with the mean equal to $|c_k|^2$. In such a case, $\sum_{k=0}^{n} X_k$ is the random norm of the generic state ψ of the system.

Proposition 9. For any $N > 1$, we have

$$P\left(\sum_{k=0}^{n} X_k < (N+1) \sum_{k=0}^{n} |c_k|^2\right) \geq \{1 - \exp[-(N+1)]\}^{n+1}$$

Proof. For any k,

$$P(|X_k - E(X_k)| \le N \,\mathrm{StDev}(X_k))$$
$$= P(|X_k - |c_k|^2| \le N|c_k|^2)$$
$$= P(0 \le X_k \le (N+1)|c_k|^2)$$
$$= \int_0^{(N+1)|c_k|^2} \exp(-y/|c_k|^2)/|c_k|^2 \, dy = 1 - \exp[-(N+1)]$$

Since the random variables $\{X_k, k = 0, 1, \ldots, n\}$ are independent, we get

$$P\left(\sum_{k=0}^{n} X_k \le (N+1)\sum_{k=0}^{n} |c_k|^2\right)$$
$$\ge P(0 \le X_k \le (N+1)|c_k|^2, k = 0, 1, \ldots, n)$$
$$= \{1 - \exp[-(N+1)]\}^{n+1}$$

Proposition 10. If the numbers $|c_k|^2$ are distinct, then

$$P\left(\sum_{k=0}^{n} X_k \le 1\right) = \sum_{k=0}^{n} \left[\prod_{\substack{j=0 \\ j \ne k}}^{n} (|c_k|^2 - |c_j|^2)\right]^{-1} [1 - \exp(-|c_k|^{-2})]|c_k|^{2n} \tag{26}$$

Proof. The random variables $\{X_k, k = 0, 1, \ldots, n\}$ are independent with the probability densities (4). Then the probability density of $X_0 + X_1$ is

$$f_0 * f_1(t) = \int_0^t f_0(x) f_1(t-x)\, dx$$
$$= \exp(-t/|c_0|^2)/(|c_1|^2 - |c_0|^2)$$
$$+ \exp(-t/|c_1|^2)/(|c_0|^2 - |c_1|^2)$$

where the asterisk means the convolution of the two probability densities. By mathematical induction we get for the probability density of $\sum_{k=0}^{n} X_k$ the expression

$$f_0 * f_1 * \cdots * f_n(t) = \sum_{k=0}^{n} \left[\prod_{\substack{j=0 \\ j \ne k}}^{n} (|c_k|^2 - |c_j|^2)\right]^{-1} \exp(-t/|c_k|^2)|c_k|^{2(n-1)} \tag{27}$$

Integrating between 0 and 1, we get (26).

7. SOME CONNECTIONS WITH WIENER AND SIEGEL'S DIFFERENTIAL SPACE

In a series of papers, Wiener and Siegel (1953, 1955) introduced statistical ensembles in a so-called "differential space," which is a Hilbert

space containing a measure for which each coordinate has an independent normal distribution with mean zero and variance equal to one. More exactly, according to Wiener and Siegel (1955), but with our notation, the differential space in the form it takes with respect to a discrete basis $\{\phi_k, k=0, 1, \ldots\}$ is an extended Hilbert space including all unnormalizable vectors as well as the normalizable vectors and such that a measure is associated with a given volume differential as follows: If $\psi=\{x_k, k=0, 1, \ldots\}$ is an arbitrary vector in this space, where the complex coordinate x_k corresponding to the vector ϕ_k may be written as

$$x_k = x_{k,1} + ix_{k,2}$$

then the weight of probability assigned to the set of points ψ contained in a small volume element

$$\prod_k dx_{k,1}\, dx_{k,2}$$

is equal to

$$\prod_k (2\pi)^{-1} \exp(-|x_k|^2/2)\, dx_{k,1}\, dx_{k,2} \tag{28}$$

There is no clear justification for introducing such a probability distribution except its nice mathematical properties. At the same time, such a model does not depend on the available data obtained at the macroscopic scale after performing some measurements on identically prepared systems. In fact, it corresponds to the probability distribution (13) when $|c_k|^2=1$ for each k.

Let U be an operator corresponding to an observable for which $\{\phi_k, k=0, 1, \ldots\}$ are the eigenfunctions and $\{u_k, k=0, 1, \ldots\}$ the corresponding eigenvalues. Let

$$\phi = \sum_k a_k \phi_k$$

be the normed state of the system, where

$$a_k = \langle \phi_k | \phi \rangle$$

Suppose that there are only a finite number, say n, of nonvanishing a_k; one can then apply the results to any extended wave function by going to the limit $n \to +\infty$ if necessary. Because the measure defined locally by (28) has been introduced without any connection to what we know about the quantum system, Wiener and Siegel invented a so-called polychotomic method in order to define a correspondence depending on ϕ between eigenvalues u_k and points ψ. Thus, they defined a functional $R(\psi, \phi)$ that may take on only the values u_k. Here ϕ is considered as fixed (the objective state of the

system) and ψ as variable (a possible state of the system). Specifically, $R(\psi, \phi)$ is to be, by definition, that eigenvalue u_k for which the ratio of the magnitudes of the associated quantities x_k of ψ and a_k of ϕ is a minimum, i.e.,

$$R(\psi, \phi) = u_j$$

if and only if

$$|x_j|^2/|a_j|^2 < |x_k|^2/|a_k|^2 \qquad \text{for all} \quad k \tag{29}$$

This method of constructing $R(\psi, \phi)$ is called the polychotomic algorithm. Wiener and Siegel proved that for any given j, the weight of the region satisfying (29) obtained as the integral of the elementary weight expression (28), over this region, is just $|a_j|^2$, which means that the probability of those ψ that satisfy (29) is $|a_j|^2$, which, according to Born, represents the quantum mechanical probability that an experiment will yield the eigenvalue u_j of the operator U. They have interpreted this result as being an explicit statistical postulate distinct from Born's statistical interpretation of the wave function, but equivalent to it as far as final results are concerned. For ease in writing we can renumber a_k so that a_j in (29) becomes a_0 and the remaining $n-1$ nonvanishing a_k receive subscripts from 1 to $n-1$ inclusive. The above result may be obtained in a more general context using the properties of the probability distributions (13) and (4).

Proposition 11. Let P be the n-dimensional direct product probability measure induced by the probability density (4). Then,

$$P\{(|x_k|^2/|c_k|^2)/|a_k|^2 > (|x_0|^2/|c_0|^2)/|a_0|^2,\ k = 1, \ldots, n-1\} = |a_0|^2$$

Proof. The event in which we are interested may be written as

$$C = \{|x_0|^2/|c_0|^2 > 0,\ |x_k|^2/|c_k|^2 > (|a_k|^2/|a_0|^2)(|x_0|^2/|c_0|^2),$$
$$k = 1, \ldots, n-1\}$$

The random variable $|x_k|^2/|c_k|^2$ is exponentially distributed with the mean 1 and the variance 1. Thus, since

$$\sum_{k=0}^{n-1} |a_k|^2 = 1$$

denoting by

$$\Upsilon(s) = \int_s^{+\infty} \exp(-y)\, dy, \qquad s > 0$$

we get

$$P(C) = \Upsilon(0)\Upsilon(|a_1|^2 y_0/|a_0|^2) \cdots \Upsilon(|a_{n-1}|^2 y_0/|a_0|^2)$$
$$= \int_0^{+\infty} \exp(-y_0/|a_0|^2)\, dy_0 = |a_0|^2$$

Thus, in our context, let us define the correspondence R as being

$$R(\psi, \phi) = u_j$$

if and only if

$$(|x_j|^2/|c_j|^2)/|a_j|^2 < (|x_k|^2/|c_k|^2)/|a_k|^2 \qquad \text{for all} \quad k \tag{30}$$

Therefore, the probability that an experiment will yield the eigenvalue u_j of the observable U, which is just the probability of the event (30), is equal to $|a_j|^2$, in accordance with Born's statistical interpretation of the wave function ϕ.

8. CONCLUSIONS

The Schrödinger equation induces a strictly deterministic evolution of the wave function of a quantum system if the wave function at the initial time is known. When the initial state of the system cannot be uniquely determined, we need at least a probability distribution on the class of possible states of the system at time t_0. At the macroscopic level, by performing measurements at time t_0 on identically prepared systems, we can get, generally, only mean values. Suppose that we know the mean probability distribution of the eigenvalues of an operator U whose system of corresponding eigenfunctions is taken as the basis of the space. Suppose that at time t_0 we know either the mean value of the observable U or, for each k, the mean value $|c_k|^2$ of the probability that the observable U takes on its eigenvalue u_k. This kind of information is not enough for a complete description of the state of the system at time t_0. Therefore, we construct, step by step, a probability measure on the set of complex coefficients of all possible, normed and nonnormed, wave functions of the system. In this construction we make use of the principle of maximum entropy from information theory taken as a mathematical tool, which allows us to determine the most random (the most unbiased) probability distribution subject to constraints given by mean values of random variables. It is quite remarkable that by applying such a strategy we obtain the exponential distribution for the square of the module of each coefficient of the wave function, the Weibull distribution for the module, and the normal distribution for the real and the imaginary parts of these coefficients. Once the space of all

possible wave functions is organized as a probability space, at instant t_0, with a Gaussian probability measure on it, the mean value of any other observable becomes a random variable whose mean value (the mean of the mean) and variance (the variance of the mean) may be determined. Also, because the Schrödinger equation generates a strictly deterministic flow in the space of possible states, the probability distribution on the state space at time t_0 will be conserved, being moved by the flow, at time t, on the transformed sets of states. Such a probabilistic approach is consistent with the general pattern met in classical statistical mechanics, but we have here a randomization process of second degree: a probability measure is constructed on the space of wave functions that are already used for making probabilistic predictions.

This approach has nothing to do with the "hidden variables" theories. It simply remakes, at a second degree of randomization, the steps learnt from the classical statistical mechanics: there is a deterministic flow on the state space, but at instant t_0 the state is unknown, and we build up the most unbiased probabilistic model on the space of possible states of the system. The only available information used in this model is expressed by mean values: the mean probabilities of the possible values of an observable U, or, when these mean probabilities are not known, the mean value of this observable. By extending the principle of maximum entropy to the complex domain as a technical mathematical tool for constructing the most random (the most unbiased, the most uniform) probability distribution subject to constraints expressed by mean values, we organize the state space as a probability space at instant t_0. This probability space is deterministically propagated in time by the flow induced by the Schrödinger equation. Luckily, the probability measure induced by this approach on the state space proves to be very manageable from the computational point of view.

ACKNOWLEDGMENT

The author gratefully acknowledges NSERC Canada research grant A-5712.

REFERENCES

Born, M. (1969). *Atomic Physics, 8th ed.*, Blakie & Son, London.

Guiasu, S. (1977). *Information Theory with Applications*, McGraw-Hill, New York.

Guiasu, S., Leblanc, R., and Reischer, C. (1982). On the principle of minimum interdependence, *Journal of Information and Optimization Sciences* **3**, 149–172.

Jaynes, E. T. (1957). Information theory and statistical mechanics, *Physical Review*, **106**, 620–630; **108**, 171–182.

Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*, Dover, New York.

Levine, R. D., and Tribus, M. (eds.) (1979). *The Maximum Entropy Formalism*, MIT Press, Cambridge, Massachusetts.

Messiah, A. (1969). *Mécanique Quantique*, Dunod, Paris.

Onicescu, O., and Guiasu, S. (1971). *Mécanique Statistique*, CISM Udine, Springer-Verlag, Heidelberg.

Smith, C. R., and Grandy, W. T. (eds.) (1985). *Maximum Entropy and Bayesian Methods in Inverse Problems*, D. Reidel, Dordrecht, Holland.

Von Neumann, J. (1932). *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag, Berlin.

Wiener, N., and Siegel, A. (1953). A new form for the statistical postulate of quantum mechanics. *Physical Review*, **91**, 1551–1560.

Wiener, N., and Siegel, A. (1955). The differential-space theory of quantum systems. *Nuovo Cimento* (*Supplemento*), **2**, 982–1003.

Wiener, N., Siegel, A., Rankin, B., and Martin, W. T. (1966). *Differential Space, Quantum Systems and Prediction*, MIT Press, Cambridge, Massachusetts.

Yeh, J. (1973). *Stochastic Processes and the Wiener Integral*, Marcel Dekker, New York.