International Journal of Theoretical Physics, Vol. 24, No. 2, 1985

Von Neumann Entropy as Information Rate

John F. Cyranski[1]

Received August 21, 1984

Recently it has been shown that quantum theory can be viewed as a classical probability theory by treating Hilbert space as a measure space $(H, B(H))$ of "events" or "hidden states." Each density operator $\hat{W}=\sum_{n=1}^{\infty} w_n \hat{\Pi}_{E_n}$ defines a set $\mathcal{M}_{\hat{W}}$ of probability measures such that $\mu(E_n)=w_n$ (all n). Coding elements $\psi \in H$ by subspaces E_n entails distortion. We show that the von Neumann entropy $S(\hat{W})=-\text{tr } \hat{W} \ln \hat{W}$ equals the effective rate at which the Hilbert space produces information with zero expected distortion, and comment on the meaning of this.

In several recent articles (Bach, 1980; Cyranski, 1982) formalisms have been developed for the treatment of quantum theory (QT) as an ordinary probability theory on the measure space $(H, B(H))$. Here H is a Hilbert space and $B(H)$ is the Borel algebra of its subsets generated using the usual topology. Bach (1980) has interpreted the elements of H as "hidden variables" whereas we (Cyranski, 1982) consider H simply as the "event space" of probability theory (Papoulis, 1965).

Since QT is usually expressed in terms of operators on the Hilbert space—or more generally, as a formalism based on a nonclassic "logic" £ defined essentially as an algebra of closed subspaces of H—it is of interest to compare the formalisms in as many ways as possible to increase our understanding of the significance of QT. In this note we consider the von Neumann entropy

$$\hat{S}(\hat{W})=-\text{Tr } \hat{W} \ln \hat{W} \tag{1}$$

and provide a novel interpretation for this quantity.

We begin by noting that if $E \in £$ (i.e., E is a closed subspace of H) then

$$q_E(\psi) \equiv \frac{\langle\psi|\hat{\Pi}_E|\psi\rangle}{\langle\psi|\psi\rangle}: H \to [0, 1] \qquad (\hat{\Pi}_E = \text{projector onto } E) \tag{2}$$

[1]Physics Department, Clark College, Atlanta, Georgia.

generalizes the set characteristic function [$\chi_\Delta(\psi)=1$ if $\psi\in\Delta$, $=0$ if $\psi\in\Delta^c$] for the vector space H and that the QT probability for $E\in\pounds$ is (Jauch, 1968) the expectation of (2):

$$P^{QT}(E)=\mathrm{Tr}\ \hat{W}\hat{\Pi}_E=\int_H d\mu(\psi)q_E(\psi), \qquad E\in\pounds \tag{3}$$

where μ is a probability measure on $B(H)$. Note that $\pounds\subset B(H)$. Each density operator $\hat{W}$ defines a partition of $\pounds$ (set of mutually orthogonal subspaces E_i) and a sequence $w_i\geq 0$ such that $\sum w_i=1$ with

$$\hat{W}=\sum w_i\hat{\Pi}_{E_i} \tag{4}$$

For a system characterized by $\hat{W}$, E_i are the subspaces containing the eigenvectors for the maximal set of commuting observables and w_i are their probabilities. Thus, $\hat{W}$ implicitly contains the totality of observationally accessible information about the system (Messiah, 1968).

Note from (3) that the relation between a probability μ on $B(H)$ and a density matrix $\hat{W}$ is many-to-one. In fact, to each $\hat{W}$ corresponds the set

$$\mathcal{M}_{\hat{W}}=\left\{\mu:B(H)\to[0,1]\,\middle|\,w_i=\int d\mu(\psi)q_{E_i}(\psi);\ i=1,2,\ldots\right\} \tag{5}$$

Let us now imagine that the "hidden" variables of the system are "communicating with" the observer over a "communication channel." While this fiction is ultimately inessential, it suggests the use of the formalism of communication theory (Berger, 1971). The "source" terminus of the channel is modeled by the "alphabet" H transmitted according to probability $\mu\in\mathcal{M}_{\hat{W}}$. The "receiver" terminus is described by the "alphabet" Y—the set of E_is defined by $\hat{W}$. Messages from the σ algebra $B(Y)$—generated by the singletons $\{E_i\}$—are received with probability $\beta(E_i)=w_i$. The channel is itself modeled by a joint probability $\bar{\mu}:B(H)\times B(Y)\to[0,1]$ having marginals $\beta(\Delta)=\bar{\mu}(H\times\Delta)$ and $\mu(\Delta)=\bar{\mu}(\Delta\times Y)$. The mutual information associated with this channel is just the "entropy"

$$I[\bar{\mu},\mu\times\beta]=\int_{H\times Y}d\bar{\mu}\ \ln\frac{d\bar{\mu}}{d(\mu\times\beta)} \tag{6}$$

The correspondence of the "hidden variables" $\psi\in H$ with the discrete set of (observable) eigenspaces $E_i\in Y$ clearly cannot be 1-1, so we generally anticipate distortion. Indeed, the formalism we employ is an effective model for data compression (Berger, 1971). In communication theory, a distortion measure $d:H\times Y\to[0,\infty]$ is introduced and one considers the minimal mutual information (6) needed to transmit messages with average distortion no greater than $D \geq 0$:

$$R(D) \equiv \inf_{\bar{\mu}} \left\{ I(\bar{\mu}, \mu \times \beta) \,\middle|\, \int d\bar{\mu}\, d \leqslant D \right\} \tag{7}$$

The function $R(D)$ is called the rate-distortion function because it is the effective rate at which the source produces information subject to the fidelity criterion, $\int d\bar{\mu}\, d \leqslant D$.

An obvious choice of distortion measure is

$$d(\psi, E_n) \equiv 1 - q_{E_n}(\psi) \tag{8}$$

which defines zero distortion if $\psi \in E_n$ and maximal distortion $(=1)$ when $\psi \perp E_n$. This generalizes the "probability of error" distortion measure (Berger, 1971): If $q_E(\psi)$ is replaced by $\chi_E(\psi)$

$$\int_H d\mu(\psi)\, d(\psi, E) = 1 - \mu(E) \tag{9}$$

is the probability that E^c is observed instead of E. The measure (8) also suggests a "pattern recognition" (or classification) "cost function": How do we assign the elements of H to the "categories" E_n and what is the "cost" of "errors"?

Now necessary and sufficient conditions for the achievement of $R(D)$ by some μ have been proved in a general context (Cyranski, 1981). For $D > 0$, $s^*(D)$ and μ, β must satisfy

$$\int_H d\mu \frac{e^{-s^* d(\psi, E_n)}}{\sum_m w_m\, e^{-s^* d(\psi, E_m)}} = 1 \qquad \forall E_n \in Y \tag{10a}$$

$$\int_H d\mu \sum_m w_m f^*(\psi, E_m)\, d(\psi, E_m) = D \tag{10b}$$

$$f^*(\psi, E_m) = \frac{e^{-s^* d(\psi, E_m)}}{\sum_n w_n\, e^{-s^* d(\psi, E_n)}} \qquad \text{a.e. } [\mu], \forall E_m \in Y \tag{10c}$$

Let $f(\psi) = d\mu / d\mu_0$ and suppose $s^* \gg 1$. Then (10a) becomes

$$1 = \int_H d\mu_0 \frac{f(\psi)\, e^{-s^* d(\psi, E_n)}}{\sum_m w_m\, e^{-s^* d(\psi, E_m)}} \approx \int_{E_n} d\mu_0 \frac{f(\psi)}{w_n} \tag{11}$$

since $\exp(-s^* d(\psi, E_k)) \approx 0$ unless $d(\psi, E_k) = 0$ and thus $\psi \in E_k$ from (8). Hence (11) and (5) tell us that $\mu \in \mathcal{M}_{\hat{W}}$ when $s^* \uparrow \infty$. From (10b) we learn

$$D = \sum_n w_n \int_H d\mu_0 f(\psi)\, d(\psi, E_n)\, e^{-s^* d(\psi, E_n)} \Big/ \left(\sum_m w_m\, e^{-s^* d(\psi, E_m)} \right)$$

$$\xrightarrow[(s^* \uparrow \infty)]{} \sum_n w_n \int_{E_n} d\mu_0 f(\psi) \frac{d(\psi, E_n)}{w_n} = 0 \tag{12}$$

so that $s^* \uparrow \infty$ corresponds to the desired case $(D \downarrow 0)$. Finally, it can be shown (Cyranski, 1981) that

$$R(D) = -\int_H d\mu_0 f(\psi) \ln\left[\sum_n w_n e^{-s^* d(\psi, E_n)}\right] - s^* D \tag{13}$$

As $s^* \uparrow \infty$ (and $D \downarrow 0$), this becomes

$$R(D) \xrightarrow[D\downarrow 0]{} -\sum_n \int_{E_n} d\mu_0 f(\psi) \ln w_n - \infty \cdot 0 = -\sum_n w_n \ln w_n = -\mathrm{Tr}\, \hat{W} \ln \hat{W} \tag{14}$$

We have thus found that the von Neumann entropy corresponds to the information rate needed to "code" the hidden variables" $\psi \in H$ into the ("observable") E_ns with zero average distortion. In other words, $S(\hat{W})$ is the source ("Nature") information rate needed to classify with perfect fidelity the "hidden variables" among the observably distinguishable mutually orthogonal subspaces defined by the QT "state" $\hat{W}$. We hope that this alternative interpretation of quantum "entropy" may prove useful in the deeper understanding of the significance of QT.

REFERENCES

Bach, A. (1980). *J. Math. Phys.*, **21**, 789–793.

Cyranski, J. F. (1982). *J. Math. Phys.*, **23**, 1074–1077.

Papoulis, A. (1965). *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York).

Werhl, A. (1978). *Rev. Mod. Phys.*, **50**, 221.

Jauch, J. M. (1968). *Foundations of Quantum Mechanics* (Addison–Wesley, Reading, Massachusetts).

Messiah, A. (1968). *Quantum Mechanics* (North-Holland, Amsterdam).

Berger, T. (1971). *Rate Distortion Theory* (Prentice-Hall, Englewood Cliffs, New Jersey).

Cyranski, J. F. (1981). *Inform. Sci.*, **24**, 217–227.