Statistical inference, distinguishability of quantum states, and quantum entanglement

V. Vedral, M. B. Plenio, K. Jacobs, and P. L. Knight
Optics Section, Blackett Laboratory, Imperial College London, London SW7 2BZ, England
(Received 23 June 1997)

We argue from the point of view of statistical inference that the quantum relative entropy is a good measure for distinguishing between two quantum states (or two classes of quantum states) described by density matrices. We extend this notion to describe the amount of entanglement between two quantum systems from a statistical point of view. Our measure is independent of the number of entangled systems and their dimensionality. [S1050-2947(97)01112-8]

PACS number(s): 03.65.Bz, 89.70.+c, 89.80.+h

Recent work has taught us that Bell's inequalities are not always a good criterion for distinguishing entangled states (i.e., those possessing a degree of quantum correlations) from disentangled states [1]. This discovery has initiated much work in quantum information theory (e.g., [2,3]) particularly concerning the search for a measure of the amount of entanglement contained within a given quantum state [4–6]. In a recent Letter [6] we presented conditions that any measure of entanglement has to satisfy. This was motivated by the fact that local actions, combined only with classical communications, should not be able to increase the amount of entanglement [4–6]. In [6] we defined our measure as the minimal distance of an entangled state to the set of disentangled states. This distance function (not necessarily a metric) could, for example, be satisfied by the quantum relative entropy (to be defined later) and by the Bures metric (for definition see, e.g., [7]). Our measure of entanglement was derived from the abstract idea of closest approximation rather than from intuitive physical grounds. In this paper we start from an entirely different point of view and derive a measure of entanglement from the idea of distinguishing two quantum states starting from classical information theory [8]. We find that these insights lead to the same measure of entanglement as in [6] (but now the quantum relative entropy is picked out from among the possible measures of ''distance''). This corroborates the results of [6] and puts them on a firm statistical basis allowing experimental tests to determine the amount of entanglement.

In order to understand our argument in the quantum case we must first describe its classical counterpart. Suppose that we are asked to distinguish between two probability distributions, taken for simplicity to be discrete. Say that we have either a fair coin with a 50-50 head-tail probability distribution or an unfair coin with 70-30 head-tail probability distribution. We are allowed to toss a *single* coin N times and we want to know which one it is. To be more general, let us say that we have a dichotomic variable with the distribution of probabilities $p(1)=p$ and $p(0)=1-p$. The probability that from N experiments (trials) we obtain n 1's and $(N-n)$ 0's is given by the binomial distribution

$$P_N(n)=\binom{N}{n}p^n(1-p)^{N-n}. \tag{1}$$

This can be written as

$$P_N(n)=\exp\{\ln\ P_N(n)\}=\exp\left\{\ln\binom{N}{n}p^n(1-p)^{N-n}\right\}. \tag{2}$$

However, using Stirling's approximation for large numbers, the exponent can be considerably simplified:

$$\ln\binom{N}{n}p^n(1-p)^{N-n}=-N\left\{\frac{n}{N}\ln\frac{n}{N}+\left(1-\frac{n}{N}\right)\ln\left(1-\frac{n}{N}\right)+\frac{n}{N}\ln p+\left(1-\frac{n}{N}\right)\ln(1-p)\right\}. \tag{3}$$

Now the quantity n/N is our *measured* frequency of 1's and likewise $1-n/N$ is the *measured* frequency of 0's in N trials. The probabilities that we infer from this distribution are given by the maximum likelihood estimate [8] $p_{inf}(1)=n/N$ and $p_{inf}(0)=1-n/N$. These are, in general, different from p and $1-p$. The crucial question we wish to ask, therefore, is: What is the probability that after N trials our inferred probabilities are q and $1-q$ if the experiment was done using a system having ''true'' probabilities p and $1-p$? In the light of the coin example we ask: What is the probability of wrongly inferring that we have a fair coin when in fact the 70-30 unfair one was used in the experiments? Clearly, the answer is given by replacing n/N by q in Eq. (3). The result in the large-N limit is

$$P_N(p\to q)=e^{-NS(q||p)}, \tag{4}$$

where

$$S(q||p):=\{q\ln q+(1-q)\ln(1-q)-q\ln p-(1-q)\ln(1-p)\} \tag{5}$$

is the so-called relative entropy, or the Kullback-Leibler distance [5,6,8,9] between the binary distributions p and q. In general, it is easy to see that the probability to confuse a distribution $\{p\}_1^M$ with $\{q\}_1^M$ in N measurements is given by

$$P_N(p\to q)=\exp\left(-N\sum_i q_i\ln q_i-q_i\ln p_i\right). \tag{6}$$

As the relative entropy is an asymmetric quantity a natural question to ask is: Why is the probability of confusing p with q different from the probability of confusing q with p? The following simple ''coin'' example will explain this. Suppose we have a fair coin and a completely unfair coin (two heads, for example). Suppose we have to decide which one it is, but we are allowed to do N experiments on only *one*, of course unknown-to-us, coin. So say we are tossing the unfair coin. Then as heads is the only possible outcome, we will never confuse the unfair coin with the fair one, as after each trial the inferred probabilities will be $p(\text{head})=1$ and $p(\text{tail})=0$. This is in fact corroborated by our formula in Eq. (6) as $e^{-\infty}=0$. On the other hand, suppose we are tossing the fair coin: Then after the first outcome, which could equally be heads or tails, we have a probability of 1/2 of confusing the coins (i.e., if the head shows up we will make the wrong inference, whereas if the tail shows up it will be the right inference). This also follows from Eq. (6) as $e^{-\ln 2}=1/2$ (note that here the formula is correct even for N small).

The central aim for us in this paper is to generalize this idea to distinguish (or, equivalently confuse) two quantum states that are completely described by their density matrices. To that end, suppose we have two states σ and ρ. How can we distinguish them? We can choose a positive operator valued measure (POVM) $\Sigma_{i=1}^{M} A_i = \mathbf{1}$ that generates two distributions via

$$p_i = \text{tr}\, A_i \sigma, \tag{7}$$

$$q_i = \text{tr}\, A_i \rho \tag{8}$$

and use classical reasoning to distinguish these two distributions. However, the choice of POVM's is not unique. It is therefore best to choose that POVM which distinguishes the distributions most, i.e., for which the relative entropy is largest. Thus we arrive at the quantity

$$S_1(\sigma||\rho) := \sup\left[A\left(\sum_i \text{tr}\, A_i\sigma \ln \text{tr}\, A_i\sigma - \text{tr}\, A_i\sigma \ln \text{tr}\, A_i\rho\right)\right],$$

where the supremum is taken over all POVM's. The above is not the most general measurement that we can make, however. In general, we have N copies of σ and ρ in the state

$$\sigma^N = \underbrace{\sigma\otimes\sigma\otimes\cdots\otimes\sigma}_{\text{total of } N \text{ terms}}, \tag{9}$$

$$\rho^N = \underbrace{\rho\otimes\rho\otimes\cdots\otimes\rho}_{\text{total of } N \text{ terms}}. \tag{10}$$

We may now apply a POVM $\Sigma_i A_i = \mathbf{1}$ acting on σ^N and ρ^N. Consequently, we define a different type of relative entropy

$$S_N(\sigma||\rho) := \sup\left[A\left(\frac{1}{N}\sum_i \text{tr}\, A_i\sigma^N \ln \text{tr}\, A_i\sigma^N - \text{tr}\, A_i\sigma^N \ln \text{tr}\, A_i\rho^N\right)\right]. \tag{11}$$

Now it can be shown that [10]

$$S(\sigma||\rho) \geqslant S_N, \tag{12}$$

where

$$S(\sigma||\rho) := \text{tr}(\sigma \ln \sigma - \sigma \ln \rho) \tag{13}$$

is the quantum relative entropy [5,6,9–12] (for a summary of the properties of quantum relative entropy see [13]). Equality is achieved in Eq. (12) if and only if σ and ρ commute [14]. However, for any σ and ρ it is true that [15]

$$S(\sigma||\rho) = \lim_{N\to\infty} S_N.$$

In fact, this limit can be achieved by projective measurements that are independent of σ [16]. From these considerations it would naturally follow that the probability of confusing two quantum states σ and ρ (after performing N measurements on ρ) is (for large N)

$$P_N(\rho\to\sigma) = e^{-NS(\sigma||\rho)}. \tag{14}$$

We would like to stress here that classical statistical reasoning applied to distinguishing quantum states leads to formula (14). There are, however, other approaches. Some take Eq. (14) for their starting point and then derive the rest of the formalism thenceforth [15]. Others, on the other hand, assume a set of axioms that are necessarily satisfied by the quantum analog of the relative entropy (e.g., it should reduce to the classical relative entropy if the density operators commute, i.e., if they are ''classical'') and then derive Eq. (14) as a consequence [10]. In any case, as we have argued here, there is a strong reason to believe that the quantum relative entropy $S(\sigma||\rho)$ plays the same role in quantum statistics as the classical relative entropy plays in classical statistics. A simple example with a ''quantum coin'' will clarify this point further [17]. Let us suppose that we have to distinguish between a pure, maximally entangled Bell state $|\phi^+\rangle=(|00\rangle+|11\rangle)/\sqrt{2}$ and a mixture $\rho=(|00\rangle\langle 00| + |11\rangle\langle 11|)/2$. Again, we have to decide which state we have by performing N experiments of our choice on it. In this case we choose to perform projections onto the state $|\phi^+\rangle=(|00\rangle+|11\rangle)/\sqrt{2}$. Then if the state ρ is in our possession, we will be successful only 50% of the time [the other 50% of the time we will obtain the orthogonal Bell state $|\phi^-\rangle=(|00\rangle-|11\rangle)/\sqrt{2}$]. So, if we perform a single experiment we have a 1/2 chance of making the wrong inference. If, on the other hand, we have $|\phi^+\rangle$, we will never confuse it with ρ since we are projecting onto the state itself that always gives a positive result. This is in direct analogy with the classical coin example and is, in addition, confirmed by Eq. (14). In general, however, the states that we have to distinguish will not be as simple as those above. Then we

would have to find the most optimal measurement to distinguish between given states in order to reproduce Eq. (14) from Eq. (11).

Now we wish to use the above reasoning to quantify entanglement. Entanglement may be understood as *the distinguishability of a given state from all entirely disentangled ones*. The question is then, in the spirit of the above discussion, as follows: What is the probability that we confuse a given state with a disentangled one after performing a total of N measurements? The less the state is entangled, the easier it is to confuse it with a disentangled one and vice versa. Thus the probability to confuse σ with a disentangled state, having performed N experiments on $\rho \in \mathcal{D}$, is of the form

$$e^{-NE(\sigma)}, \tag{15}$$

where $E(\sigma)$ is the entanglement (obviously, if $E=0$, then the state is indistinguishable from a disentangled one since it is disentangled itself). In comparison with Eq. (14), we define $E(\sigma)$ to be

$$E(\sigma) := \min_{\rho \in \mathcal{D}} S(\sigma||\rho), \tag{16}$$

where $\mathcal{D}$ is the set of all disentangled states. So for the entanglement of σ we use the quantum relative entropy with that disentangled ρ which is the most *indistinguishable* from σ. Obviously, the greater the entanglement of a state, the smaller the chance of confusing it with a disentangled state in N measurements. Note that Eq. (16) is the same measure as that suggested in our previous Letter [6]. There we showed that the Bures metric, when used instead of $S(\sigma||\rho)$, would also be a good measure of entanglement. However, the Bures distance is a symmetric quantity and arises from different statistical consideration from those used above (see [7] for an overview). Thus, depending on the way we decide to make our measurements, we obtain different ways of comparing the results (i.e., different ''distances'' between probability distributions), which in turn determine our entanglement measure [more correctly, the quantity that is to replace $S(\sigma||\rho)$ in Eq. (16)]. The convention that we use here assumes performing measurements on ρ. We could, of course, envisage making measurements on σ, in which case our measure of entanglement would be $E(\sigma) := \min_{\rho \in \mathcal{D}} S(\rho||\sigma)$. However, for σ being, for example, a maximally entangled Bell state, this quantity would be infinite. This agrees with our statistical interpretation that a Bell state, when measurements are performed on it, could never be confused with a disentangled state and Eq. (15) gives probability zero of confusion. However, in order to avoid dealing with physically undesired infinite amount of entanglement we keep to the convention given in Eq. (16).

We see that the above treatment does not refer to the number (or indeed dimensionality) of the entangled systems. This is a desired property as it makes our measure of entanglement universal. However, in order to perform minimization in Eq. (16) we need to be able to define what we mean by a disentangled state of, say, N particles. As pointed out in [6], we believe that this can be done inductively. Namely, for two quantum systems A_1 and A_2 we define a disentangled state as one that can be written as a convex sum of disentangled states of A_1 and A_2 as [3,6,18]

$$\rho_{12} = \sum_i p_i \, \rho_i^{A_1} \otimes \rho_i^{A_2}, \tag{17}$$

where $\Sigma_i p_i = 1$ and the p's are all positive. Now, for N entangled systems $A_1, A_2, \ldots, A_N$, the disentangled state is

$$\rho_{12\cdots N} = \sum_{\mathrm{perm}\{i_1 i_2 \cdots i_N\}} r_{i_1 i_2 \cdots i_N} \rho^{A_{i_1} A_{i_2} \cdots A_{i_n}} \otimes \rho^{A_{i_{n+1}} A_{i_{n+2}} \cdots A_{i_N}}, \tag{18}$$

where $\Sigma_{\mathrm{perm}\{i_1 i_2 \cdots i_N\}} r_{i_1 i_2 \cdots i_N} = 1$, all r's are positive, and $\Sigma_{\mathrm{perm}\{i_1 i_2 \cdots i_N\}}$ is a sum over all possible permutations of the set of indices $\{1,2,\ldots,N\}$. To clarify this let us see how this looks for four systems

$$\begin{aligned} \rho_{1234} = \sum_i & \, p_i \, \rho_i^{A_1 A_2 A_3} \otimes \rho_i^{A_4} + q_i \, \rho_i^{A_1 A_2 A_4} \otimes \rho_i^{A_3} + r_i \, \rho_i^{A_1 A_3 A_4} \\ & \otimes \rho_i^{A_2} + s_i \, \rho_i^{A_2 A_3 A_4} \otimes \rho_i^{A_1} + t_i \, \rho_i^{A_1 A_2} \otimes \rho_i^{A_3 A_4} + u_i \, \rho_i^{A_1 A_3} \\ & \otimes \rho_i^{A_2 A_4} + v_i \, \rho_i^{A_1 A_4} \otimes \rho_i^{A_2 A_3}, \end{aligned} \tag{19}$$

where, as usual, all the probabilities $p_i, q_i, \ldots, v_i$ are positive and add up to unity. Equations (18) and (19), at least in principle, define the disentangled states for any number of entangled systems. In practice, unfortunately, this might still not be enough to minimize the relative entropy to obtain the amount of entanglement. So far a good criterion for decomposition into the above form exists for two particles only, when either both are spin 1/2 or one is spin 1/2 and the other one is spin 1 [3,18] (however, some progress has been made by Horodecki [19]). The above definition of a disentangled state is justified by extending the idea that local actions cannot increase the entanglement between two quantum systems [4–6]. In the case of N particles we have N parties (Alice, Bob, Charlie, ..., Wayne) all acting locally on their systems. The general action that also includes communications can be written as [6]

$$\rho \rightarrow \sum_{i_1, i_2, \ldots, I_N} A_{i_1} \otimes B_{i_2} \otimes \cdots \otimes W_{i_N} \, \rho \, A_{i_1}^\dagger \otimes B_{i_2}^\dagger \otimes \cdots \otimes W_{i_N}^\dagger \tag{20}$$

and it can be easily seen that this action does not alter the form of a disentangled state in Eqs. (18) and (19). In fact, Eq. (18) is the most general state invariant *in form* under the transformation given by Eq. (20). We suggest this as a definition of a disentangled state for $N \geqslant 3$, i.e., it is the most general state invariant in form under local POVM and classical communications. This definition of N-particle entanglement means that we say that we do not have N-particle entanglement even if subsets of the N particles are individually entangled. We define it this way so that it answers the question, are all N particles entangled, rather than the question, is there any entanglement at all between the particles. If we wanted to answer the latter question, then clearly the definition of a disentangled N-particle state would be one that could be written as

$$\rho=\sum_i p_i\rho_i^A\otimes\rho_i^B\otimes\cdots\otimes\rho_i^W. \tag{21}$$

We have in this work derived our previously proposed measure of entanglement from an entirely different perspective. The amount of entanglement is now seen as the quantity that determines ''the least number of measurements that is needed to distinguish a given state from a disentangled one.'' This therefore strengthens the argument for using Eq. (16) as a universal measure of entanglement. In addition, it opens up the possibility both to understand the meaning of entanglement from a different, more operational, point of view and to measure the amount of entanglement for more than two quantum systems.

We thank A. Ekert and C. A. Fuchs for discussions and useful comments on this subject. This work was supported by the European Union, the United Kingdom Engineering and Physical Sciences Research Council, a Feodor-Lynen grant of the Alexander von Humboldt Foundation, the British Council, the New Zealand Vice Chancellors' Committee, and the Knight Trust.

[1] N. Gisin, Phys. Lett. A **210**, 151 (1996), and references therein; A. Peres, Phys. Rev. A **54**, 2685 (1996); M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **223**, 1 (1996).

[2] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996); D. Deutsch, A. Ekert, R. Josza, C. Macchiavello, S. Popescu, and A. Sanpera, Phys. Rev. Lett. **77**, 2818 (1996).

[3] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. Lett. **78**, 574 (1997).

[4] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[5] V. Vedral, M. A. Rippin, and M. B. Plenio, J. Mod. Opt. (to be published).

[6] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Phys. Rev. Lett. **78**, 2275 (1997).

[7] C. Fuchs and C. M. Caves, LANL e-print quant-ph/9604001.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, New York, 1991).

[9] A. Wehrl, Rev. Mod. Phys. **50**, 221 (1978).

[10] M. J. Donald, Commun. Math. Phys. **105**, 13 (1986); Math. Proc. Camb. Philos. Soc. **101**, 363 (1987).

[11] G. Lindblad, Commun. Math. Phys. **40**, 147 (1975).

[12] G. Lindblad, Commun. Math. Phys. **39**, 111 (1974).

[13] M. Ohya, Rep. Math. Phys. **27**, 19 (1989).

[14] C. A. Fuchs, Ph.D. thesis, The University of New Mexico, 1996 (unpublished); LANL e-print quant-ph/9601020.

[15] F. Hiai and D. Petz, Commun. Math. Phys. **143**, 99 (1991).

[16] M. Hayashi, LANL e-print quant-ph/9704040.

[17] A. Ekert (private communication).

[18] R. Horodecki and M. Horodecki, Phys. Rev. A **54**, 1838 (1996).

[19] P. Horodecki, LANL e-print quant-ph/9703004.