Towards a proof of two conjectures from quantum inference concerning quantum limits to knowledge of states

1991 J. Phys. A: Math. Gen. 24 L415

(http://iopscience.iop.org/0305-4470/24/8/005)

J. Phys. A: Math. Gen. 24 (1991) L415–L419. Printed in the UK

LETTER TO THE EDITOR

Towards a proof of two conjectures from quantum inference concerning quantum limits to knowledge of states

Kingsley R W Jones

School of Physics, University of Melbourne, Parkville 3052, Melbourne, Australia

Received 10 December 1990

Abstract. We give a new entropic analogue of the recently reported information theoretic limits to knowledge of states. A natural relationship between the quantum correlation information and the quantum mechanical entropy is thereby revealed. In addition we make some progress towards a rigorous proof of both results and a complete solution to the problem of asymptotic optimal measurement. In particular we employ elementary convex analysis to prove that the optimal operator valued measure must be a rank-one projection valued measure.

From the researches of numerous workers, most notably Davies and Lewis [1], Helstrom [2] and Holevo [3] we now know that information about quantum states is most generally obtained via a statistical correlation of form

$$p(\hat{A}|\psi) = \langle\psi|\hat{A}|\psi\rangle \tag{1}$$

where ψ is the state of the system and $\hat{A}$ denotes the apparatus reading, being a positive Hermitian operator belonging to some positive operator valued measure [1, 2, 3] (hereafter a POV measure). Note that closure of the conditional viz $\int p(\hat{A}|\psi)\,\mathrm{d}\Pi[\hat{A}] = 1$, requires the POV measure to satisfy a possibly non-orthogonal (overcomplete) resolution of unity, $I = \int \hat{A}\,\mathrm{d}\Pi[\hat{A}]$, where $\mathrm{d}\Pi[\hat{A}]$ denotes an infinitesimal positive Hermitian operator, the corpus of which comprises the operator valued measure. Note that this extended formalism is of physical value in that examples of such generalized measurements certainly exist [4].

This letter concerns recent developments that have to do with the inversion of such data for the express purpose of inferring the quantum state of an ensemble. This is the quantum problem of determining initial conditions. Our interest lies in developing general constraints upon how well this can be done.

Similar investigations have been made before using what we might term the quantum analogue of standard statistical decision theory [2, 3]. However, in statistics there is an alternative methodology, the often maligned procedure of Bayesian inference. The first serious study of this possibility within quantum mechanics is due to Wootters [5]. Recently I have extensively elaborated this approach [6, 7] and shown how it leads naturally to a theory of quantum inference that provides both straightforward data inversion and a simple general approach to the problem of constraining knowledge of states [7]. Here we shall briefly review the basic idea and then go on to continue earlier work [8] aimed at developing the fundamental quantum limits for finite-dimensional Hilbert space.

0305-4470/91/080415+05$03.50

Consider N observations, with possibly different POV measures, upon N members of a pure ensemble described by state ψ. Let the observed data be denoted $\Phi_N \equiv \{\hat{A}_k\}_{k=1}^N$. Then the conditional probability of this event is equal to

$$p(\Phi_N|\psi) = \prod_{k=1}^{N} \langle\psi|\hat{A}_k|\psi\rangle. \tag{2}$$

A natural inversion of this density to obtain $p(\psi|\Phi_N)$ results when we employ Bayes's rule and normalize the right-hand side by the quantity

$$p(\Phi_N) = \int p(\Phi_N|\psi)\, \mathrm{d}\hat{\Omega}_{\hat{\psi}} \tag{3}$$

where $\mathrm{d}\hat{\Omega}_{\hat{\psi}}$ is a prior density chosen as the unique invariant measure on ray space. This procedure we justified at length in [7].

The natural quantum estimator of the state ψ is the inferred density matrix

$$\rho(\Phi_N) = \int |\psi\rangle\langle\psi|\, p(\psi|\Phi_N)\, \mathrm{d}\hat{\Omega}_{\hat{\psi}} \tag{4}$$

for which an equation in closed form was first given in [7]. Notice that $\rho(\Phi_N)$ is always mixed reflecting imperfect inference of the unknown pure state ψ. This observation will form one plank of our development.

The second plank concerns a quantity that measures the efficacy of any set $\mathscr{A}_N$ of N POV measures thought of as an N-trial apparatus for the measurement of quantum states. This is the quantum correlation information of the apparatus [7]

$$\{\psi, \Phi_N\}[\mathscr{A}_N] \equiv \iint p(\Phi_N) p(\psi|\Phi_N) \log p(\psi|\Phi_N)\, \mathrm{d}\Phi_N\, \mathrm{d}\hat{\Omega}_{\hat{\psi}} \tag{5}$$

where $\mathrm{d}\Phi_N \equiv \Pi_{k=1}^N \mathrm{d}\Pi_k[\hat{A}_k]$. This quantity originates from Shannon's communication theory [9] but indeed has application to any problem of Bayesian inference (a largely overlooked possibility). In this context it gives the average information in nats (1 nat = 1/log 2 bits) that readings Φ_N of $\mathscr{A}_N$ give about the unknown state ψ.

The third plank concerns the fact that any N-trial inferred density is of generic form

$$p(\psi|\Phi_N) = \frac{1}{p(\Phi_N)} \prod_{k=1}^{N} \langle\psi|\hat{A}_k|\psi\rangle. \tag{6}$$

It is then intuitively obvious that the confidence of any such density is N-constrained by the Hilbert space geometry. We now give two inequalities that measure that constraint.

Since all N-trial inferred densities $p(\psi|\Phi_N)$ are built from N factors $\langle\psi|\hat{A}|\psi\rangle$, each of non-zero support in Hilbert space we have the two fundamental quantum inequalities:

$$\{\psi, \Phi_N\}[\mathscr{A}_N] \leq \sup_{\Phi_N} \int p(\psi|\Phi_N) \log p(\psi|\Phi_N)\, \mathrm{d}\hat{\Omega}_{\hat{\psi}} \tag{7}$$

$$S[\rho(\Phi_N)] \geq \inf_{\Phi_N} S[\rho(\Phi_N)] \tag{8}$$

where $S[\rho] = -\mathrm{Tr}\, \rho \log \rho$. Hereafter, we fix the Hilbert space as being complex of finite dimension d. Existence of a finite upper bound in the first inequality was demonstrated in [7]. However, the second is new, its existence follows trivially from the fact that $S[\rho] \in [0, \log d]$.

The non-trivial aspect of these inequalities lies in the solution of a variational problem over all possible apparatus $\mathscr{A}_N$ so as to determine the supremum and infimum as a function of N. The problem is made tractable by observing that we need only seek the sup and inf over all possible choices of N non-negative Hermitian operators $\hat{A}_k$, $k \in [1, N]$. *That is to say we know the sup and inf are attained as some inferred* density of some apparatus, but we also know that all such densities are parametrized by N such operators.

In the article [8] we asserted a solution to the variational problem associated with the first inequality, restricted to the case of rank-one projection valued measures only. We now make a stronger assertion.

In both cases the fundamental bounds are set by the following inferred density

$$p(\psi|\Phi_N) = \frac{(N+d-1)!}{(d-1)!\,N!}(|\langle\psi|\phi\rangle|^2)^N \tag{9}$$

where the choice of ϕ is unimportant. This we call the inferred density of maximal information. Furthermore, this remains the solution when the restriction to rank-one projection valued measures is relaxed. That is to say, the above density is extremal by both criteria for all POV measures.

Here we shall prove that the rank-one projection valued measures are indeed optimal on all POV measures. What we have yet to prove is the conjecture that the above density solves the resulting restricted variational problem.

Contingent upon the unproven component of our assertion are the following two conjectures

$$\sup_{\Phi_N} \int p(\psi|\Phi_N) \log p(\psi|\Phi_N)\, \mathrm{d}\hat{\Omega}_{\psi} = \log\left[\frac{(N+d-1)!}{(d-1)!\,N!}\right] - \sum_{k=1}^{d-1} \frac{N}{N+k} \tag{10}$$

$$\inf_{\Phi_N} S[\rho(\Phi_N)] = \log(N+d) - \frac{N+1}{N+d}\log(N+1). \tag{11}$$

These are obtained by substituting the density (9) into the right-hand sides of equations (7) and (8). For suitable computational methods see [8]. For instance, one finds that the optimal inferred density matrix is

$$\rho(\phi) = \frac{1}{N+d}\mathbf{1} + \frac{N}{N+d}|\phi\rangle\langle\phi|.$$

Interestingly, this approaches the pure state $|\phi\rangle\langle\phi|$ in the limit $N\to\infty$.

The physical interpretation of both inequalities is very simple. In the absence of *a priori* knowledge one cannot know more than the quantity (10) nats of information about the finite degrees of freedom of the state of a pure ensemble containing N members. Nor can the inferred density matrix have an entropy that is less than the lower bound (11). We stress that these limits are absolutely fundamental and cannot be circumvented.

Denote by $\mathscr{P}_\phi$ the set of all inferred densities $p(\psi|\Phi_N)$ defined by N rank-one projectors and by $\mathscr{P}_{\hat{A}}$ its more general cousin upon N positive Hermitian operators. Note that in general any family of unconstrained probability densities form a convex set. This fact leads to a simple proof based upon elements of convex analysis [10]. For a closely related problem see Davies [11].

Consider any convex $\cup$ (concave $\cap$) real valued functional defined upon a convex domain. It is an elementary result of convexity theory that the maximum (minimum) of such a convex (concave) functional must be attained at an extremal point of the convex domain (the result is not of itself powerful enough to say which of such points). Furthermore, for any subdomain that is not convex, one finds that the required extrema are attained upon those points of the subdomain that are also extremal points of its convex hull [10].

One now checks that the right-hand sides of (7) and (8) define a convex and a concave functional, respectively. Next one verifies that any inferred density belonging to $\mathscr{P}_{\hat{A}}$ can be written as a convex combination of at most d^N members of $\mathscr{P}_\phi$. This point requires some attention.

One obtains a canonical decomposition into just this number by replacing each $\hat{A}_k$ in (6) by its spectral decomposition (the operators are positive and Hermitian). The only point of difficulty involves splitting up the contributions to the normalization $p(\Phi_N)$ so as to ensure that the combination coefficients sum to one. This is expedited by recognizing a freedom to renormalize the $\hat{A}_k$ so as to have unit trace. We note in passing that this actually means POV measures can be identified up to equivalence classes labelled by *density matrix valued* measures. The use of such measures is to be preferred.

This brief excursion shows that the set of densities $\mathscr{P}_{\hat{A}}$ contains the set $\mathscr{P}_\phi$ as its extremal points. Note that this result requires that $\mathscr{P}_\phi$ include only densities derived from *rank-one* projectors. Now it is important to recognize that neither set of densities is convex because some convex combinations of their elements cannot be written in the form (6). However, it is now obvious that the extremal points of the *convex* set conv $\mathscr{P}_\phi$ are just the full set of points $\mathscr{P}_\phi$. Furthermore, this convex hull contains $\mathscr{P}_{\hat{A}}$ as a proper subset.

Applying the theorem described at the outset we can now deduce the desired result that the extrema of both functionals are attained upon densities derived from the rank-one projection valued measures.

We have reduced the problem of constraining knowledge of quantum states to the examination of rank-one POV measures only. To this restricted problem we already have good evidence that (9) provides the solution. We now give a rigorous proof of (10) and (11) for the special case $N=2$.

Using the formulae of [7, 8] one can verify that the expressions to be maximized and minimized (respectively) in equations (7) and (8) are

$$\text{RHS} = \log d(d+1) + 2[\Psi(2) - \Psi(d+2)] - \log(1 + \operatorname{Tr} \hat{A}_1\hat{A}_2) + \frac{2 \operatorname{Tr} \hat{A}_1\hat{A}_2}{1 + \operatorname{Tr} \hat{A}_1\hat{A}_2} \tag{12}$$

$$\text{RHS} = S\left[\frac{1}{d+2}\mathbb{1} + \frac{2}{d+2}\,\frac{\hat{A}_1 + \hat{A}_1\hat{A}_2 + \hat{A}_2\hat{A}_1 + \hat{A}_2}{[2 + 2\operatorname{Tr} \hat{A}_1\hat{A}_2]}\right] \tag{13}$$

where we have utilized the freedom to normalize all $\hat{A}_k$ to have unit trace and Ψ denotes the digamma function.

Now if we set $p = \operatorname{Tr} \hat{A}_1\hat{A}_2$ in the first case and recognize that $p \in [0, 1]$ it becomes a simple calculus problem to verify that the maximum is attained when $\hat{A}_1$ and $\hat{A}_2$ are equal and idempotent.

In the second case we see that the inferred density matrix assumes the generic form

$$\rho(\hat{A}_1, \hat{A}_2) = \frac{1}{d+2}\mathbb{1} + \frac{2}{d+2}\rho^*(\hat{A}_1, \hat{A}_2)$$

where $\rho^*(\hat{A}_1, \hat{A}_2)$ is itself a density matrix. Solving the appropriate variational problem under this assumption alone shows that the minimum is attained uniquely when ρ^* is idempotent. Inspection of the formula

$$\rho^*(\hat{A}_1, \hat{A}_2) = \frac{(\hat{A}_1+\hat{A}_2)^2 + (\hat{A}_1 - \hat{A}_1^2) + (\hat{A}_2 - \hat{A}_2^2)}{[2 + 2\,\mathrm{Tr}\,\hat{A}_1\hat{A}_2]}$$

then reveals that this is only possible when $\hat{A}_1$ and $\hat{A}_2$ are again equal and idempotent.

So we certainly know (9) is the optimal inference for $N=2$, the same method of proof appears to extend to embrace all N in the second case. Using the explicit formula for $\rho(\Phi_N)$ that was given in [7] we can show that the inferred density matrix has the general decomposition

$$\rho(\Phi_N) = \frac{1}{N+d}\,1 + \frac{N}{N+d}\,\rho^*(\Phi_N)$$

where $\rho^*(\Phi_N)$ is certainly of trace one. If one can show that it is also positive (which should be so) then the result (11) follows without difficulty.

The bounds (10) and (11) would be of little value were it not for the fact that there exist apparatus $\mathscr{A}_N$ whose average behaviour approaches the optimal possible inference arbitrarily closely in the limit as $N \to \infty$.

One such apparatus was characterized in the paper [8], it corresponds to repeated use of the *isotropic* POV measure $p(\phi|\psi) = |\langle\phi|\psi\rangle|^2$ where ϕ varies continuously over the pure state manifold and is ascribed a uniform measure. Holevo [3] has studied such POV measures before as an example of what he calls a covariant instrument.

Having calculated its asymptotic behaviour in [8] we now strongly suspect that this measure is uniquely the optimal apparatus. To prove this one must show that this measure attains uniquely the supremum of $\{\psi, \Phi_N\}[\mathscr{A}_N]$, for all N (we already know that it is asymptotically optimal). Proving the lesser statement concerning the optimal *inference* establishes the absolute constraints, whereas the harder optimal *apparatus* problem establishes their feasibility.

References

[1] Davies E B and Lewis J T 1970 *Commun. Math. Phys.* **17** 239
[2] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[3] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[4] Arthurs E and Kelly J L Jr 1965 *Bell Syst. Tech. J.* **44** 725
She C Y and Heffner H 1966 *Phys. Rev.* **152** 1103
[5] Wootters W K 1980 *PhD thesis* Center for Theoretical Physics, University of Texas at Austin
[6] Jones K R W 1990 *PhD thesis* H H Wills Physics Laboratory, University of Bristol
[7] Jones K R W 1990 *Ann. Phys.* in press
[8] Jones K R W 1991 *J. Phys. A: Math. Gen.* **24** 121–130
[9] Shannon C E and Weaver W 1949 *The Mathematical Theory of Communication* (Urbana: Illinois Press)
[10] Rockafellar R T 1970 *Convex Analysis* (Princeton: Princeton University Press) p 342
[11] Davies E B 1978 *IEEE Trans. Inform. Theory* **IT-24** 596