ASYMPTOTICS IN QUANTUM STATISTICS

RICHARD D. GILL

University of Utrecht and Eurandom

Observations or measurements taken of a quantum system (a small number of fundamental particles) are inherently random. If the state of the system depends on unknown parameters, then the distribution of the outcome depends on these parameters too, and statistical inference problems result. Often one has a choice of what measurement to take, corresponding to different experimental set-ups or settings of measurement apparatus. This leads to a design problem—which measurement is best for a given statistical problem. This paper gives an introduction to this field in the most simple of settings, that of estimating the state of a spin-half particle given n independent copies of the particle. We show how in some cases asymptotically optimal measurements can be constructed. Other cases present interesting open problems, connected to the fact that for some models, quantum Fisher information is in some sense non-additive. In physical terms, we have non-locality without entanglement.

AMS subject classifications: 62F12, 62P35.
Keywords and phrases: quantum statistics, information, spin half, qubit, two level system.

1 Introduction

The fields of quantum statistics and quantum probability have a reputation for being esoteric. However, in our opinion, quantum mechanics, from which they surely derive, is a fascinating source of probabilistic and statistical models, unjustly little known to 'ordinary' statisticians and probabilists.

Quantum mechanics has two main ingredients: one deterministic, one random. In isolation from the outside world a quantum system evolves deterministically according to Schrödinger's equation. That is to say, it is described by a state or wave-function whose time evolution is the (reversible) solution of a differential equation. On the other hand, when this system comes into interaction with the outside world, as when for instance measurements are made of it (photons are counted by a photo-detector, tracks of particles observed in a cloud chamber, etc.) something random and irreversible takes place. The state of the system makes a random jump and the outside world contains a record of the jump. From the state of the system at the time of the interaction one can read off, according to certain rules, the probability distribution of the macroscopic outcomes and the new state of the system. See Penrose (1994) for an eloquent discussion of why there is something paradoxical in the peaceful coexistence of these two principles;

see Percival (1998) for interesting stochastic modifications to Schrödinger's equation which might offer some reconciliation.[1]

Till recently, most predictions made from quantum theory involved such large numbers of particles that the law of large numbers takes over and predictions are deterministic. However, technology is rapidly advancing to the situation that really small quantum systems can be manipulated and measured (e.g., a single ion in a vacuum-chamber or a small number of photons transmitted through an optical communication system). Then the outcomes definitely are random. The fields of quantum computing, quantum communication, and quantum cryptography are rapidly developing and depend on the ability to manipulate individual quantum systems. Especially of interest are assemblages of two-level systems, known as qubits or spin half systems in quantum computing and information. Theory and conjecture are much further than experiment and technology, but the latter are following steadily.

In this paper we will introduce as simply as possible the model of quantum statistics and consider the problem of how to best measure the state of an unknown spin-half system. We will survey some recent results, in particular, from joint work with O.E. Barndorff-Nielsen and with S. Massar (Barndorff-Nielsen and Gill, 2000; Gill and Massar, 2000). This work has been concerned with the problem posed by Peres and Wootters (1991): can more information be obtained about the common state of n identical quantum systems from a single measurement on the joint system formed by bringing the n systems together, or does it suffice to combine separate measurements on the separate systems? A useful tool for our studies is the quantum Cramér-Rao bound with its companion notion of quantum information, introduced by Helstrom (1967). Actually there are several ways to define quantum information with different resulting Cramér-Rao type bounds (Yuen and Lax, 1973, Stratonovich, 1973, Belavkin, 1976, Holevo, 1982).

Quantum statistics mainly consists of exact results in various rather special models; see the books of Helstrom (1976) and Holevo (1982). Just as in ordinary statistics, the Cramér-Rao bound on the variance of an unbiased estimator is rarely achieved exactly (only in so-called quantum exponential models). In any case, one would not want in practice to restrict attention to unbiased estimators only. There are results on optimal invariant methods, but again, not many models have the structure that these results are applicable and even then the restriction to invariant statistical methods is not entirely compelling.

One might hope that asymptotically it would be possible to achieve the

[1] Also highly recommended: Sheldon Goldstein, 'Quantum mechanics without observers', Physics Today, March, April 1998; letters to the editor, Physics Today, February 1999.

Cramér-Rao bound. However, asymptotic theory is so far very little developed in the theory of quantum statistics, one reason being that the powerful modern tools of asymptotic statistics (contiguity, local asymptotic normality, and so on) are just not available[2] since even if we are considering measurements of n identical quantum systems, there is no a priori reason to suppose that a particular sequence of measurements on n quantum systems together will satisfy these conditions. Here, we make a little progress through use of the van Trees inequality (see Gill and Levit, 1995), a Bayesian Cramér-Rao bound, which will allow us to make asymptotic optimality statements without assuming or proving local asymptotic normality. Another useful ingredient will be the recent derivation of the quantum Cramér-Rao bound by Braunstein and Caves (1994), linking quantum information to classical expected Fisher information in a particularly neat way.

We will show that, for certain problems, a new Cramér-Rao type inequality of Gill and Massar (2000) does provide an asymptotically achievable bound to the quality of an estimator of unknown parameters. For some other problems the issue remains largely open and we identify situations where Peres and Wootter's question has an affirmative answer: there can be appreciably more information in a joint measurement of several particles than in combining separate measurements on separate particles. This clarifies an earlier affirmative answer of Massar and Popescu (1995), which turned out only for small samples to improve on separate measurements. It also clarifies the recent findings of Vidal et al. (1998).

Helstrom wrote in the epilogue to his (1976) book: "*Mathematical statisticians are concerned with asymptotic properties of estimators. When the parameters of a quantum density operator are estimated on the basis of many independent observations, how does the accuracy of the estimates depend on the number of the observations as that number grows very large? Under what conditions have the estimators asymptotically normal distributions? Problems such as these, and still others that doubtless will occur to physicists and mathematicians, remain to be solved within the framework of the quantum mechanical theory.*" More than twenty years later this programme is still hardly touched (some of the few contributions are by Brody and Hughston, 1998, and earlier papers, and Holevo, 1983) but we feel we have made a start here.

In $20 \pm \epsilon$ pages (even when $\pm\epsilon = +12$) it is difficult to give a complete introduction to the topic, as well as a clear picture of recent results. The classic books by Helstrom and Holevo mentioned above are still the only books on quantum statistics and they are very difficult indeed to read for a beginner. A useful resource is the survey paper by Malley and Hornstein (1993). However the latter authors, among many distinguished writers both

[2] though R. Rebolledo is working on a notion of quantum contiguity

from physics and from mathematics, take the stance that the randomness occurring in quantum physics cannot be caught in a standard Kolmogorovian framework. We argue elsewhere (Gill, 1998), in a critique of an otherwise excellent introduction to the related field of quantum probability (Kümmerer and Maassen, 1998), that depending on what you mean by such a statement, this is at best misleading, and at worst simply untrue. With more space at our disposal we would have included extensive worked examples; however they have been replaced by *exercises* so that the reader can supply some of the extra pages (but—unless you are Willem van Zwet—leave the starred exercises for later).

Some references which we found specially useful in getting to grips with the mathematical modelling of quantum phenomena are the books by Peres (1995), and Isham (1995). To get into quantum probability, we recommend Biane (1995) or Meyer (1986). Also highly recommended are the lecture notes of Preskill (1997), Werner (1997) and Holevo (1999).

This introductory section continues with three subsections summarizing the basic theory: first the mathematical model of states and measurements; secondly the basic facts about the most simple model, namely of a two-state system; and thirdly the basic quantum Cramér-Rao bound. That third subsection finishes with a glimpse of how one might do asymptotically optimal estimation in one-parameter models: in a preliminary stage obtain a rough estimate of the parameter from a small number of our n particles. Estimate the so-called quantum score at this point, and then go on to measure it in the second stage on the remaining particles. Section 2 states a recent new version of the quantum Cramér-Rao bound which makes precise how one might trade information between different components of a parameter vector. Section 3 outlines the procedure for asymptotically optimal estimation of more than one parameter, again a two-stage procedure. This is work 'in progress', so some results are conjectural, imprecise, or improvable. In a final short section we try to explain how some of our results are connected to the strange phenomenon of *non-locality without entanglement*, a hot topic in the theory of quantum information and computation.

1.1 The basic set-up

Quantum statistics has two basic building blocks: the mathematical specification of the state of a quantum system, to be denoted by $\rho = \rho(\theta)$ as it possibly depends on an unknown parameter θ, and the mathematical specification of the measurement, denoted by M, to be carried out on that system. We will give the recipe for the probability distribution of the observable outcome (a value x of a random variable X say) when measurement M is carried out on a system in state ρ. Since the state ρ depends on an unknown parameter θ, the distribution of X depends on θ too, thereby setting a statistical problem of how best to estimate or test the value of θ. Since we may in practice have a choice of which measurement M to take, we have a design problem of choosing the best measurement for our purposes. There is also a recipe for the state of the system after measurement, depending on the outcome, and depending on some further specification of the measurement; see Preskill (1997), Werner (1997), Bennett et al. (1998), or Holevo (1999). We do not need it here.

For simplicity we restrict attention to finite-dimensional quantum systems. The state of a d-dimensional quantum system can be summarized or specified by a $d \times d$ complex matrix ρ called the density matrix of the system. For instance, when we measure the spin of an electron in a particular direction only two different values can occur, conventionally called 'up' and 'down'. This is just one example of a two level system, requiring a $d =$ two-dimensional state space for its description. Similarly if we measure whether a photon is polarized in a particular direction by passing it through a polarization filter, it either passes or does not pass the filter. Again, polarization measurements on a single photon can be discussed in terms of a two-dimensional system. If we consider the spins of n electrons, then 2^n different outcomes are possible and the system of n electrons together (or rather, their spins) is described by a $d \times d$ matrix ρ with $d = 2^n$. The future quantum computer might consist of an assemblage of n atoms at very low temperature, each of which could be found in its ground state or in an excited state; interacting together they have a 2^n-dimensional state space.

Definition 1.1 (Density matrix) *The density matrix ρ of a d-dimensional quantum system is a $d \times d$ self-adjoint, nonnegative matrix of trace 1.*

'Self-adjoint' means that $\rho^* = \rho$ where $*$ denotes the complex conjugate and transpose of the matrix. That ρ is nonnegative means that $\psi^* \rho \psi \geq 0$ for all column vectors ψ (since ρ is self-adjoint this quadratic form is a real number). We often use the Dirac bra-ket notation whereby $|\psi\rangle$ (called a ket) is written for the complex column vector ψ and $\langle\psi|$ (a bra) is written for its adjoint, the row vector containing the complex conjugates of its elements. The quadratic form $\psi^* \rho \psi$ is then written $\langle \psi \mid \rho \mid \psi \rangle$. The notation allows one to graphically distinguish numbers from matrices, as in $\langle \psi \mid \psi \rangle$ versus $|\psi\rangle \langle\psi|$, and to specify vectors through labels or descriptions as in $|\,\text{lable}\,\rangle$ or $|\,\text{description}\,\rangle$

The diagonal elements of a density matrix must be nonnegative reals adding up to one. Moreover by the eigenvalue-eigenvector decomposition of self-adjoint matrices we can write $\rho = \sum_i p_i \, |i\rangle \langle i|$ where the kets $|i\rangle$ are orthonormal eigenvectors of ρ, $\langle i \mid j \rangle = \delta_{ij}$, and the p_i are the eigenvalues: nonnegative real numbers adding up to one. More correctly one should speak of the eigenvalue-eigenspace decomposition; the eigenvectors corresponding to a particular eigenvalue are an arbitrary orthonormal basis of the eigenspace. Even if the eigenvalue is simple (the eigenspace one-dimensional), the eigenvector is only uniquely defined up to multiplication by an arbitrary complex number of modulus one.

One says that the density matrix ρ represents the *mixed* state obtained by taking with probability p_i the system in the *pure* state described by the density matrix $|i\rangle\langle i|$. A pure state $|\psi\rangle\langle\psi|$ is often referred to through its *state vector* $|\psi\rangle$, also called the wave-function. One may multiply $|\psi\rangle$ by an arbitrary complex number of modulus one, and it still represents the same state.

Definition 1.2 (Measurement) *A measurement M on a d-dimensional quantum system taking values x in a measurable space $(\mathcal{X}, \mathcal{A})$ is described by an* operator-valued probability measure (oprom), *that is, a collection of self-adjoint matrices $M(A) : A \in \mathcal{A}$ such that*

1. *$M(\mathcal{X}) = \mathbf{1}$, the identity matrix,*

2. *Each $M(A)$ is non-negative.*

3. *For disjoint A_i, $M(\cup_i A_i) = \sum_i M(A_i)$.*

Note that these three rules are the ordinary axioms of a probability measure on $(\mathcal{X}, \mathcal{A})$, except that the measure takes values in the self-adjoint matrices instead of the real numbers. The sample space $\mathcal{X}$ might be the real numbers or a subset thereof, with the Borel sigma algebra, but it could also be anything else.

Now we can give the so-called *trace-rule* telling us the probability distribution of the random outcome X when M is used to measure ρ:

Definition 1.3 (Trace rule) *The probability distribution of the outcome X is given by*

$$\text{(1)} \qquad \Pr\{X \in A\} = \operatorname{trace}(\rho M(A)), \qquad A \in \mathcal{A}$$

Exercise 1.1 (Legitimacy of trace rule) *Prove that* (1) *indeeds defines a probability measure on $\mathcal{X}, \mathcal{A}$.*

One can argue from basic principles of quantum mechanics that however one measures a quantum system, the result must be an affine mapping from density matrices to the space of probability distributions on the outcome space. It is a theorem that any such mapping can be represented by an oprom. Thus the class of oproms contains all conceivable measurements. On the other hand, as we will see later, any oprom can be realised by some concrete experimental set-up, at least in principle, so the definition captures exactly what it should. The inaccurate abbreviations 'povm' (positive operator valued measure) and 'pom' (probability operator matrices) are widespread.

Oproms are often called *generalised measurements*, to contrast them with a special subclass of measurements called *simple measurements*, defined as follows:

Definition 1.4 (Simple measurement) *A simple measurement Π on a d-dimensional quantum system is a real-valued measurement such that each $\Pi(A)$ is idempotent, i.e., is an orthogonal projector onto a subspace of $\mathbb{C}^d$.*

One can now show that the measurement takes on at most d different values, i.e., there exist $x_1, \ldots, x_k \in \mathcal{X} = \mathbb{R}$ with $k \leq d$ such that $\Pi(\{x_1, \ldots, x_k\}) = \mathbf{1}$. With $\Pi(x_i)$ as abbreviation for $\Pi(\{x_i\})$, the matrices $\Pi(x_i)$ project onto k orthogonal subspaces of $\mathbb{C}^d$ together spanning the whole space. Define a self-adjoint matrix $\hat{X}$ (to be distinguished from the random variable X representing the outcome of the measurement) by $X = \sum_i x_i \Pi(x_i)$. Then the x_i are the eigenvalues of $\hat{X}$ and the $\Pi(x_i)$ project onto the eigenspaces. Conversely, given a self-adjoint matrix $\hat{X}$ one can construct a corresponding simple measurement or projector-valued probability measure. In this role we call $\hat{X}$ an *observable*. It follows that the expected value of the outcome of a measurement of $\hat{X}$ is given by $\text{trace}(\rho \hat{X})$. For an ordinary real function f (e.g., square, inverse, logarithm, ...) one defines the same function of the observable $\hat{X}$ by $f(\hat{X}) = \sum_i f(x_i)\Pi(x_i)$, and the expected value of the outcome of a measurement of the observable $f(\hat{X})$ is $\text{trace}(\rho f(\hat{X}))$.

Simple measurements are often called von Neumann measurements. We will occasionally use the term 'proprom' (projector-valued probability measure). Physicists generally agree that any simple measurement could in principle be implemented in practice.

'Between measurements' a quantum system evolves deterministically according to the Schrödinger equation, a differential equation for the component pure states $|i\rangle$ of a given mixed system. One thinks of a measurement as taking place instantaneously. After the measurement, the quantum system jumps to a new state (depending on the outcome x); this is called 'the collapse of the wave function'. Some simple rules specify what happens, but we will not give them here.

If we bring two separate quantum systems together into some kind of interaction, then their future evolutions will be linked together. Measurements can be made on the 'joint system', including all the separate measurements on each of the separate systems but many more besides. Mathematically this is modelled as follows:

Definition 1.5 (Product system) *Consider two quantum systems, of dimension d and d', in states ρ and ρ' respectively. Together the two are in the state $\rho \otimes \rho'$ in $\mathbb{C}^d \otimes \mathbb{C}^{d'} = \mathbb{C}^{d \times d'}$ where $\otimes$ denotes tensor product (of matrices, vectors, or spaces as appropriate).*

For the reader who is not familiar with tensor products, the tensor product of $\mathbb{C}^d$ with $\mathbb{C}^{d'}$ has as basis the tensor product of each element of a basis of $\mathbb{C}^d$ with each element of a basis of $\mathbb{C}^{d'}$. One can take linear combinations of tensor products $\psi \otimes \psi'$ by expanding bilinearly in chosen bases of the two spaces. Tensor products of matrices are defined in the natural way by how they operate on products of vectors: $X \otimes X' \psi \otimes \psi' = X\psi \otimes X'\psi'$. The trace of a tensor product of two matrices is the product of the traces.

Suppose M and M' are measurements on two separate quantum systems ρ and ρ'. Then we can define a joint measurement $M \otimes M'$ on the combined system in the obvious way, taking values in the product of the outcome spaces of M and M'.

Exercise 1.2 (Product measurement) *Show that the outcome of measurement of $M \otimes M'$ on a system in state $\rho \otimes \rho'$ is distributed as independent realisations of measurement of M and M' on ρ and ρ' respectively.*

However, the important point is that bringing two quantum systems together allows many more measurements than just product measurements (which, as we saw from exercise 1.2, are not very interesting).

Product systems are important for two main reasons. Firstly, one of the main themes of this paper is going to be: if we have n independent systems each in the same state $\rho(\theta)$ (i.e., in identical states all depending on the *same* unknown parameter θ), can we learn more about θ from a joint measurement on the d^n dimensional combined system $\rho^{\otimes n}(\theta)$? In the next section we will discuss some of the history and other background to this question, which has been the subject of a series of papers in recent years. Secondly, product systems play a role in the realisation of generalised measurements. It is a theorem (due to Naimark) that *any generalised measurement whatever can be realised by a simple measurement after a 'quantum randomisation'.* That is to say, given any measurement M there exists a so-called ancillary system and a fixed and known state ρ' of that system, and a simple measurement Π on the joint system $\rho \otimes \rho'$ such that $\mathrm{trace}(\rho M(A)) = \mathrm{trace}(\rho \otimes \rho' \,\Pi(A))$ for all A and ρ.

1.2 Spin half

In order to make the above rather abstract concepts a little more concrete, let us go to the most simple special case, $d = 2$, called a spin-half system, a two-level system, or a qubit. We will see that we can associate the state of a spin-half system with a real vector $\vec{a}$ of length less than or equal to 1 in ordinary three dimensional space, and a simple measurement—which can take on at most two different values—with a direction in space, or a unit vector $\vec{u}$. The trace rule (1) reduces to a simple formula involving $\vec{a}$ and $\vec{u}$. The model applies to the famous Stern-Gerlach experiment, featuring in many introductory textbooks on quantum physics. In that experiment silver atoms were made to pass through a magnetic field with large gradient in a certain direction. Each atom was either deflected upwards or downwards with respect to this direction. The deflection is due to the spin of the outermost electron in the silver atom, which can be characterized by a vector $\vec{a}$. The orientation of the magnet determines which measurement is being taken, i.e., the value of $\vec{u}$.

First we take some time to study some special features of the 2 × 2 self-adjoint matrices. The properties we find will greatly simplify calculations. Define the *Pauli spin matrices* as follows:

Definition 1.6 (Pauli spin matrices)

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2}$$

These three matrices are self adjoint, each has trace zero and determinant minus one, hence each has eigenvalues ±1. They satisfy (check this yourself!)

$$\begin{aligned} \sigma_x\sigma_y = -\sigma_y\sigma_x &= i\sigma_z, \\ \sigma_y\sigma_z = -\sigma_z\sigma_y &= i\sigma_x, \\ \sigma_z\sigma_x = -\sigma_x\sigma_z &= i\sigma_y, \\ \sigma_x^2 = \sigma_y^2 = \sigma_z^2 &= \mathbf{1}. \end{aligned} \tag{3}$$

An arbitrary self-adjoint 2 × 2 complex matrix has to be of the form

$$X = \begin{pmatrix} u+z & x-iy \\ x+iy & u-z \end{pmatrix} \tag{4}$$

where x, y, z, u are uniquely determined real numbers. Thus we can write

$$X = u\mathbf{1} + x\sigma_x + y\sigma_y + z\sigma_z. \tag{5}$$

Specializing to density matrices, the requirement that $\operatorname{trace}\rho = 1$ imposes the condition that $u = \frac{1}{2}$. The requirement that ρ is nonnegative is equivalent to its determinant being nonnegative, or $u^2 - z^2 - x^2 - y^2 \geq 0$, or $x^2 + y^2 + z^2 \leq \frac{1}{2}^2$. It is convenient to write

$$\rho = \rho(\vec{a}) = \frac{1}{2}(\mathbf{1} + \vec{a} \cdot \vec{\sigma}) \tag{6}$$

where $\vec{a} = (a_x, a_y, a_z) \in \mathbb{R}^3$ and satisfies

$$\|\vec{a}\|^2 = a_x^2 + a_y^2 + a_z^2 \leq 1 \tag{7}$$

while $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, a vector of matrices, and $\cdot$ denotes the inner-product. Thus the space of density matrices of a two-dimensional quantum system can be represented by the closed unit ball B in three-dimensional Euclidean space. The sphere S, or surface of the unit ball, corresponds to density matrices $\frac{1}{2}(\mathbf{1}+\vec{a}\cdot\vec{\sigma})$ with $\|\vec{a}\|^2 = 1$, which are singular since their determinant is zero. Such a density matrix has therefore eigenvalues 0 and 1 and represents a pure state.

The density matrix of a pure state is a projector matrix, projecting onto a one-dimensional subspace of $\mathbb{C}^2$. Letting $\vec{u}$ denote a unit vector in $\mathbb{R}^3$, let us write $\Pi(\vec{u}) = \rho(\vec{u}) = \frac{1}{2}(\mathbf{1} + \vec{u}\cdot\vec{\sigma})$ for this matrix. We can also write $\Pi(\vec{u}) = |\vec{u}\rangle\,\langle\vec{u}|$ for a certain state vector $|\vec{u}\rangle$ which we will determine in a moment. Check using (3) that $\Pi(\vec{u})$ is idempotent, and that $\Pi(\vec{u})$ and $\Pi(-\vec{u})$ commute (in fact, their product is the zero matrix) and add to the identity matrix. Thus the projectors $\Pi(\vec{u})$ and $\Pi(-\vec{u})$ project onto two orthogonal one-dimensional subspaces of $\mathbb{C}^2$, the spaces generated by $|\vec{u}\rangle$ and $|-\vec{u}\rangle$ respectively. The only other projector matrices are $\mathbf{0}$ and $\mathbf{1}$, projecting onto the zero subspace and the whole space of $\mathbb{C}^2$ respectively.

Given a density matrix $\rho = \rho(\vec{a})$, define the unit vector $\vec{u} = \vec{a}/\|\vec{a}\|$ and probabilities $\alpha = \frac{1}{2}(1 + \|\vec{a}\|)$, $\beta = \frac{1}{2}(1 - \|\vec{a}\|)$. If $\|\vec{a}\| = 0$, $\vec{u}$ can be chosen arbitrarily. Then we can write

$$\rho(\vec{a}) = \alpha\rho(\vec{u}) + \beta\rho(-\vec{u}). \tag{8}$$

This shows that $\rho(\vec{a})$ has eigenvalues α and β, and its eigenvectors are $|\vec{u}\rangle$ and $|-\vec{u}\rangle$ respectively. One may consider the state $\rho(\vec{a})$ as the mixture, with probabilities α and β, of the pure states with state vectors $|\vec{u}\rangle$ and $|-\vec{u}\rangle$ (though this is only one of many representations of ρ as a mixture of pure states).

The vector $\vec{u}$ is a point on the unit sphere in $\mathbb{R}^3$. Let θ and ϕ denote its polar coordinates, where $\theta \in [0, \pi]$ is the latitude measured from the North pole (z-axis) and $\phi \in [0, 2\pi)$ is the longitude, measured from the x-axis. (We should really say co-latitude rather than latitude). Thus $\vec{u} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. We are going to show that $|\vec{u}\rangle = |\theta, \phi\rangle$, the vector in $\mathbb{C}^2$ defined by

$$|\theta, \phi\rangle = \begin{pmatrix} e^{-i\phi/2}\cos(\theta/2) \\ e^{i\phi/2}\sin(\theta/2) \end{pmatrix}. \tag{9}$$

Note that $\langle\theta,\phi \mid \theta,\phi\rangle = 1$ while

(10)

$$\begin{aligned}
|\theta,\phi\rangle\langle\theta,\phi| &= \begin{pmatrix} e^{-i\phi/2}\cos(\theta/2) \\ e^{i\phi/2}\sin(\theta/2) \end{pmatrix} \begin{pmatrix} e^{i\phi/2}\cos(\theta/2) & e^{-i\phi/2}\sin(\theta/2) \end{pmatrix} \\
&= \begin{pmatrix} \cos^2(\theta/2) & e^{-i\phi}\cos(\theta/2)\sin(\theta/2) \\ e^{i\phi}\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) \end{pmatrix} \\
&= \tfrac{1}{2}\begin{pmatrix} 1+\cos(\theta) & (\cos\phi - i\sin\phi)\sin\theta \\ (\cos\phi + i\sin\phi)\sin\theta & 1-\cos\theta \end{pmatrix} \\
&= \tfrac{1}{2}(1+\vec{u}\cdot\vec{\sigma}) = \Pi(\vec{u}).
\end{aligned}$$

Any complex vector $|\xi\rangle$ of length 1 can be written as $e^{i\alpha}|\theta,\phi\rangle$ for some $\alpha \in [0,2\pi)$ and polar coordinates θ,ϕ and $|\xi\rangle\langle\xi|$ is the same for all α. We summarize our findings so far:

Rule 1.1 (Spin-half density matrices, projectors) *The density matrix $\rho(\vec{a})$, where $\vec{a}$ is a point in the unit ball in $\mathbb{R}^3$, has eigenvalues $\frac{1}{2}(1\pm\|\vec{a}\|)$ and and normalized eigenvectors $|\vec{u}\rangle = |\theta,\phi\rangle$, $|-\vec{u}\rangle = |\pi-\theta,\phi+\pi\rangle$, where θ and ϕ are the polar coordinates of $\vec{u} = \vec{a}/\|\vec{a}\|$. The projector matrix $\Pi(\vec{u})$ projects onto the one-dimensional subspace of $\mathbb{C}^2$ spanned by $|\vec{u}\rangle$. The projector onto the space orthogonal to this, spanned by $|-\vec{u}\rangle$, is $\Pi(-\vec{u})$.*

Let $\vec{u}$ and $\vec{v}$ be two unit vectors in $\mathbb{R}^3$. Since $\Pi(\vec{u}) = |\vec{u}\rangle\langle\vec{u}|$ we see that $\operatorname{trace}\Pi(\vec{u})\Pi(\vec{v}) = \operatorname{trace}|\vec{u}\rangle\langle\vec{u}|\,|\vec{v}\rangle\langle\vec{v}| = \langle\vec{v}\mid\vec{u}\rangle\langle\vec{u}\mid\vec{v}\rangle = |\langle\vec{u}\mid\vec{v}\rangle|^2$. On the other hand, using the properties (3) of the Pauli matrices, one readily computes $\operatorname{trace}\Pi(\vec{u})\Pi(\vec{v}) = \frac{1}{2}(1+\vec{u}\cdot\vec{v})$. Now $\vec{u}\cdot\vec{v}$ is the cosine of the angle between the vectors $\vec{u}$ and $\vec{v}$, hence $\frac{1}{2}(1+\vec{u}\cdot\vec{v})$ is the squared cosine of half the angle between $\vec{u}$ and $\vec{v}$.

Rule 1.2 (Calculation rule) *The absolute value of the squared inner product between the complex vectors $|\vec{u}\rangle$ and $|\vec{v}\rangle$ in $\mathbb{C}^2$ is the squared cosine of half the angle between the corresponding unit vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$. In particular, opposite points on the unit sphere correspond to orthogonal vectors in $\mathbb{C}^2$.*

We can now describe the probability distributions of all *simple* measurements of a spin-half system.

The state of the system is modelled by a 2 × 2 density matrix of the form $\rho(\vec{a}) = \frac{1}{2}(1+\vec{a}\cdot\vec{\sigma})$ where $\vec{a}$ is a point in the closed unit ball in $\mathbb{R}^3$.

The non-trivial simple measurements take on just two different values. Consider a simple measurement $M = \Pi$ taking values in a set $\mathcal{X}$ consisting of just two elements, let's call these elements ±1. The measurement is determined by the two projectors $\Pi(\pm)$, which should project onto orthogonal one-dimensional subspaces of $\mathbb{C}^2$. Each subspace is generated by a vector of the form $|\vec{u}\rangle$ for some $\vec{u}$ on the unit sphere, and the associated projectors are $\Pi(\vec{u})$. Recall that opposite points $\pm\vec{u}$ on the unit sphere correspond to orthogonal vectors $|\pm\vec{u}\rangle$ in $\mathbb{C}^2$, and hence to orthogonal projectors $\Pi(\pm\vec{u})$. Thus a projector-valued probability measure for a simple measurement with values in $\mathcal{X}$ is given by $M(\pm 1) = \Pi(\pm\vec{u}) = \frac{1}{2}(\mathbf{1} \pm \vec{u}\cdot\vec{\sigma})$ for some $\vec{u}$.

We apply the trace rule (1) to compute the probabilities of the two outcomes ±1 when the simple measurement $M(\pm 1) = \Pi(\pm\vec{u})$ is carried out on a system in the state $\rho(\vec{a}) = \frac{1}{2}(\mathbf{1} + \vec{a}\cdot\vec{\sigma})$. Extending Rule 1.2, the reader should verify that these probabilities are

$$\text{trace}\,\rho(\vec{a})\Pi(\pm\vec{u}) = \frac{1}{2}(1 \pm \vec{a}\cdot\vec{u}). \tag{11}$$

Using further rules for the state of the system after measurement, it turns out that after measurement the system is in the pure state $\rho(\pm\vec{u})$ according to the outcome ±1. One can therefore go on to compute probabilities of the series of outcomes of a series of simple measurements carried out on one particle.

In the Stern-Gerlach experiment, the initial state of the silver atom is described by the density matrix $\rho(\vec{0}) = \frac{1}{2}\mathbf{1}$. One can think of this state as corresponding to an electron having spin in a random direction $\vec{u}$ uniformly distributed over the unit sphere. Indeed, if one takes the mean of $\rho(\vec{u}) = \frac{1}{2}(\mathbf{1} + \vec{u}\cdot\vec{\sigma})$ with $\vec{u}$ uniformly distributed over the sphere, the matrix $\frac{1}{2}\mathbf{1}$ results, though this representation of the 'completely random' state $\rho(\vec{0})$ as a mixture of pure states is not unique.

Exercise 1.3 (A generalised measurement of spin-half system) *Let $M(A) = \int_A \Pi(\vec{u})\mathrm{d}\vec{u}/2\pi$ where $\mathrm{d}\vec{u}$ denotes integration with respect to Lebesgue surface measure on S. Show that M is a generalized measurement on a spin-half system with values in S, and compute the distribution of the outcome of this measurement on the system $\rho(\vec{a})$. This measurement would be physically realised by somehow coupling the spin-half system with a particle moving on the sphere and measuring the position of that particle.*

Exercise 1.4 (A generalized measurement of n spin-half systems*) *For the state-space $(\mathbb{C}^2)^{\otimes n}$ define $|\vec{u}\rangle_n = |\vec{u}\rangle \otimes \cdots \otimes |\vec{u}\rangle$ and define $\Pi_n(\vec{u}) = |\vec{u}\rangle_n\langle\vec{u}|_n$. Define $M(A) = (n+1)\int_A \Pi_n(\vec{u})\mathrm{d}\vec{u}/4\pi$ and show that $M(S)$ is the projector onto the $n+1$ dimensional subspace of vectors, invariant under permutation of the n components of $(\mathbb{C}^2)^{\otimes n}$. Call this subspace S_n and note that $\text{trace}\,\rho^{\otimes n}\Pi_{S_n} = 1$. Show that M defines a generalized measurement on n identical copies of a spin-half system with values in S, and compute the distribution of the outcome of this measurement on the system $\rho(\vec{v})$.*

A Stern-Gerlach magnet oriented in the direction $\vec{u}$ implements the simple measurement $M(\pm 1) = \Pi(\pm\vec{u})$. Since for $\vec{a} = \vec{0}$ the probabilities (11) both equal $\frac{1}{2}$, one will find electrons with spin in the directions $\pm\vec{u}$ with equal probabilities. Electrons in the emerging '+' beam are in the pure state $\rho(\vec{u})$. Sending them through a Stern-Gerlach device with orientation $\vec{v}$ splits them again, now with probabilities $\frac{1}{2}(1\pm\vec{u}\cdot\vec{v})$ (the squared cosine of half the angle between the directions $\vec{u}$ and $\vec{v}$) into two beams of electrons in the states $\rho(\pm\vec{v})$, and so on.

If the electrons started out in the arbitrary mixed state $\rho(\vec{a})$, then the first Stern-Gerlach magnet splits them into two output beams in the pure states $\rho(\pm\vec{u})$ in the proportions $\frac{1}{2}(1\pm\vec{a}\cdot\vec{u})$. So, if $\vec{a}$ was unknown, we do learn something about it from counting the numbers of electrons in each beam. Further operations on the output beams however will not teach us any more as the state of the electrons in either output beam no longer depends on $\vec{a}$.

If we are allowed to measure a large number of electrons each in the same mixed state $\rho(\vec{a})$, we see that a large number of Stern-Gerlach measurements in three linearly independent directions will enable us to determine $\vec{a}$. The question we will study in the rest of the paper is: what is the best way to do this? Will it suffice to use simple measurements on separate particles or can we do better by using more sophisticated measurements, in particular, joint measurements on several particles simultaneously?

One can consider rotating a given coordinate system in $\mathbb{R}^3$ in such a way as to transform the vectors $\vec{a}$ and $\vec{u}$ representing a state or a simple measurement into convenient choices, e.g., we will in the future claim that 'without loss of generality $\vec{a} = (0,0,a_3)$' which makes $\rho(\vec{a})$ a diagonal matrix. How to do this is given by the following (more difficult) exercise:

Exercise 1.5 (Rotation of coordinate system*) *For given unit-vector $\vec{u}$ and angle θ define $U = \exp(-i\theta\vec{u}\cdot\vec{\sigma}/2)$. Then $UU^* = U^*U = \mathbf{1}$, i.e., U is a unitary transformation of $\mathbb{C}^2$, and $U\rho(\vec{a})U^* = \rho(\vec{b})$ where $\vec{b}\in\mathbb{R}^3$ results from $\vec{a}$ by rotation about $\vec{u}$ through an angle θ.*

This result belongs to the *representation theory of groups*, an important topic having deep connections with quantum theory. It is a curious fact that with $\theta = 2\pi$ the operator U is equal to $-\mathbf{1}$. Though, in this case, U works on a density matrix by a rotation through 360°, it does not transform a state vector to itself but to its negative. A rotation through 720° or the angle 4π is needed to transform a vector to itself. The fact that one complete revolution does not transform a vector into itself has experimentally observable consequences, namely certain interference effects.

1.3 Quantum Cramér-Rao inequality

Consider a quantum statistical model whereby the density matrix ρ depends on an unknown parameter θ. Possibly θ is a vector but we will not emphasize that fact in the notation. In particular, a spin-half system has a density matrix $\rho = \rho(\vec{a})$ depending on a point $\vec{a}$ in the closed unit ball B. By S we denote the unit sphere, the boundary of B. Interesting statistical models could therefore have a one-, two- or three-dimensional parameter θ, specifying a curve, a surface, or an open region of B. Of particular interest are one- and two-dimensional *pure-state models*, specifying a curve in S and the whole of S respectively. Results are strikingly different according to whether the true value of θ corresponds to a point in S or in the interior of B. By a *mixed-state model* we mean a model in the interior of B. By the *full model*, pure or mixed, we mean the model: 'ρ is in S', and 'ρ is in the interior of B' respectively. By the natural parametrization of these models we mean the parametrization $\rho = \rho(\vec{u})$, $\rho = \rho(\vec{a})$ respectively.[3]

The quantum Cramér-Rao bound involves a collection of self-adjoint matrices λ_i called the quantum score matrices, one for each component of θ, and a quantum information matrix.These are defined as follows.

Definition 1.7 (Quantum score matrices) *Suppose $\rho = \rho(\theta)$ depends on parameters $\theta = (\theta_1, \ldots, \theta_k)$. Suppose that ρ is differentiable with respect to θ. Define the quantum scores $\lambda_i = \lambda_i(\theta)$ as the self-adjoint matrices which solve the equations*

$$\rho_i = \frac{\partial \rho(\theta)}{\partial \theta_i} = \frac{1}{2}(\lambda_i \rho + \rho \lambda_i). \tag{12}$$

When θ is one-dimensional, we drop the index and write λ for the quantum score. Note that the $\lambda_i = \lambda_i(\theta)$ typically depends on θ. The quantum score is also known as the symmetric logarithmic derivative of ρ with respect to θ_i. If ρ and its derivative ρ_i with respect to θ_i commute, then λ_i is nothing else than the derivative of $\log \rho$. By using a basis of $\mathbb{C}^d$ making ρ diagonal, $\rho = \sum p_j \left|j\right\rangle \left\langle j\right|$, one can solve (12) to obtain

$$\langle j \mid \lambda_i \mid j' \rangle = \frac{2 \langle j \mid \rho_i \mid j' \rangle}{p_j + p_{j'}}. \tag{13}$$

If some p_j are zero, the corresponding elements of λ_i may be chosen arbitrarily (subject to self-adjointness) without effect on subsequent calculations. If ρ is a pure state, then $\rho^2 = \rho$ and it follows from differentiating this equation with respect to θ_i that in this case $\lambda_i = 2\rho_i$.

Exercise 1.6 (Mean quantum score zero) *Show that the quantum score has expectation zero, that is, the distribution of a measurement of the observable λ_i has mean zero, or* $\operatorname{trace}(\rho\lambda_i) = 0$.

[3] It would be nice to express conditions and results in the language of differential geometry, i.e., independent of the specific parametrizations of the models under consideration.

Exercise 1.7 (Spin half, mixed) *Consider the full mixed-state spin-half model $d = 2$, $\rho = \frac{1}{2}(1 + \vec{\theta} \cdot \vec{\sigma})$, where θ is three-dimensional and satisfies $\sum_i \theta_i^2 < 1$. Then $\rho_i = \sigma_i$ for each i. At the point $\theta = (0, 0, \xi)$ the density matrix is diagonal with diagonal elements $\frac{1}{2}(1 \pm \xi)$ and the quantum scores are found from* (13) *to be σ_x, σ_y and $(1 - \xi)^{-2}(-\xi 1 + \sigma_z)$.*

Exercise 1.8 (Spin half, pure) *The full pure-state spin-half model has everything as in the previous exercise but now with $\sum_i \theta_i^2 = 1$. A two-dimensional parametrization is called for, using, e.g., the polar coordinates of the unit vector $\vec{\theta}$. However, on the Northern hemisphere we can stick to $\theta = (\theta_1, \theta_2)$ with $\theta_3 = +(1 - \theta_1^2 - \theta_2^2)^{1/2}$ and we find that at $\theta = (0, 0)$ the quantum scores are σ_x and σ_y.*

Exercise 1.9 (n copies) *Suppose $\rho^{(n)}(\theta) = \rho^{\otimes n}(\theta)$. Then the quantum scores are given by*

$$\lambda_i^{(1)} \otimes 1 \otimes \cdots \otimes 1 + \ldots + 1 \otimes \cdots \otimes 1 \otimes \lambda_i^{(1)}. \tag{14}$$

Now we can define the quantum information matrix and state the quantum Cramér-Rao bound.

Definition 1.8 (Quantum information matrix) *The quantum information matrix I_Q is defined by*

$$(I_Q)_{ii'} = \frac{1}{2} \operatorname{trace}(\rho(\lambda_i \lambda_{i'} + \lambda_{i'} \lambda_i)). \tag{15}$$

Check that this defines a real, positive semi-definite matrix!

Exercise 1.10 (n copies, continued) *Show from* (14) *and Exercise 1.6 that the quantum information $I_Q^{(n)}$ for a parameter θ in the system $\rho^{\otimes n}$ is just n times the quantum information for θ in a single copy of the system.*

Theorem 1.1 (Quantum Cramér-Rao bound) *Define $I_M(\theta)$ to be the Fisher information matrix for the parameter θ in the distribution of the outcome of a measurement M on the quantum system $\rho(\theta)$. Then (with respect to the usual ordering of symmetric positive semi-definite matrices) we have $I_M(\theta) \leq I_Q(\theta)$.*

The result in this form was proved by Braunstein and Caves (1994) for a one-dimensional parameter, but the general result is an easy consequence of this by considering the information for arbitrary linear combinations; see also Barndorff-Nielsen and Gill (2000). As a corollary one obtains Helstrom's original form of the theorem: the variance of an unbiased estimator of θ

based on the outcome of an arbitrary measurement M can not be smaller than $I_Q(\theta)^{-1}$.

Both Braunstein and Caves' (1994) and Helstrom's (1967) proofs involve little more than the Cauchy-Schwarz inequality for the complex inner-product trace X^*Y between two self-adjoint matrices. And just as in the usual proof of the Cramér-Rao inequality, as a by-product one can read off from the proof necessary and sufficient conditions for equality to hold. For a one-dimensional parameter, an attractive *sufficient* condition is that M should be a simple measurement of the observable λ:

Exercise 1.11 (Optimal M for 1-d θ) *Show for one-dimensional θ that if M is the simple measurement of the observable λ, i.e., its values are in one-to-one correspondence with the eigenvalues of λ and each $M(x)$ is the projection onto the corresponding eigenspace, then $I_M = I_Q$.*

Life would be easy (but less interesting) if this provided a practical solution. However, typically λ will depend on θ, and typically in such a strong way that the eigenspaces of λ (and not just eigenvalues) depend on θ. Thus the best measurement of θ in terms of Fisher information depends on the true value of θ. However things are very simple in the following example:

Exercise 1.12 *Suppose all $\rho(\theta)$ commute, i.e., have common eigenspaces. Show that the $\rho_i(\theta)$ then also commute for all i and θ. Show that a simple measurement of the common eigenspaces of all these matrices has Fisher information equal to the quantum information for all values of θ.*

In this example, we have $\rho = \sum_i p_i(\theta)\,|i\rangle\,\langle i|$, a mixture with mixing distribution depending on θ over the fixed pure states $|i\rangle$. The optimal measurement asks which of these pure states the system is in, in other words it is represented by the projector-valued probability measure with elements $|i\rangle\,\langle i|$. This measurement has outcome i with probability $p_i(\theta)$. The quantum information matrix is the Fisher information matrix for this distribution.

The result of Exercise 1.11 does give hope for a clear solution to the problem of estimating a one-dimensional parameter, for large n, for the system $\rho^{\otimes n}(\vec{a}(\theta))$, as was first pointed out by Barndorff-Nielsen and Gill (2000). Suppose the parameter θ is identified, i.e., the mapping from θ to $\rho(\theta)$ is one to one. Then there are a finite number of simple measurements, the distributions of whose outcomes identify θ. For example, in the spin-half case $\rho = \frac{1}{2}(1 + \vec{\theta} \cdot \vec{\sigma})$, measurements of σ_1, σ_2, σ_3 result in Bernoulli trials with probabilities $\frac{1}{2}(1 \pm \theta_i)$. Suppose that from consistent estimators of the $a_i(\theta)$ we can construct a consistent estimator of θ. Use a growing number but vanishing proportion of copies of our quantum system with which to 'pre-estimate' θ consistently. Call this preliminary estimator $\widetilde{\theta}$. Next, compute the quantum score for θ at $\widetilde{\theta}$, determine its eigenspaces, and implement the corresponding simple measurement on all remaining copies of the system. This gives us an i.i.d. sample from some distribution $p(\cdot|\theta;\widetilde{\theta})$. Estimate θ by maximum likelihood on these observations conditional on the observed value of $\widetilde{\theta}$. The result $\widehat{\theta}$ will be an estimator approximately normally distributed about θ with variance approximately $1/nI(\theta;\widetilde{\theta})$ where $I(\theta;\widetilde{\theta})$ is the Fisher information for θ in one of these observations, given $\widetilde{\theta}$. Now for n large we have arranged that $\widetilde{\theta}$ is close to the true value of θ. We may hope that the eigenspaces of $\lambda(\widetilde{\theta})$ are close to the eigenspaces of $\lambda(\theta)$ and hence that the Fisher information in one observation (one simple measurement) of $\lambda(\widetilde{\theta})$ is close to that in one observation of $\lambda(\theta)$. But the latter achieves the quantum Cramér-Rao bound at θ. Thus under suitable smoothness conditions $I(\theta;\widetilde{\theta})$ will be close to $I_Q(\theta)$ and hence the asymptotic distribution of our final estimator close to normal about θ with variance $1/nI_Q(\theta)$. This is coming close to saying that $\widehat{\theta}$ is asymptotically optimal.

We know that no unbiased estimator of θ can have smaller variance. However that does not tell us no estimator whatever can do better, e.g., in terms of mean square error. Indeed the phenomenon of super-efficiency is just as present here as in ordinary statistics. In order to make a compelling optimality statement about our estimator we must either restrict attention to a sub-class of nicely behaved estimators, or make optimality statements which are of a Bayesian or a minimax nature. A very useful tool, which can be used in any of these approaches, is the van Trees inequality which says for a one-dimensional parameter θ with prior distribution $\pi(\mathrm{d}\theta)$, under some regularity conditions, that the expected (with respect to the prior) mean square error of a completely arbitrary estimator of θ is bounded by one divided by the expected Fisher information for the parameter plus the information, with respect to location, in the prior distribution. This writer prefers to restrict the class of estimators according to some regularity condition. We will go into this in more detail in the next section, but before that, let us consider the multiparameter case. We will see that a more fundamental complication arises: at a fixed parameter value, quantum scores for different components of the parameter may not commute.

Exercise 1.13 (Quantum information for spin-half models) *In exercises 1.8 and 1.7 we noted the score matrices for the full pure-state model $\rho = \rho(\vec{u})$ and for the full mixed-state model $\rho = \rho(\vec{a})$. Show that, at $\vec{u} = (0,0,1)$ in the first case and at $\vec{a} = (0,0,\xi)$ in the second case, the quantum information matrices for $\theta = (u_1,u_2)$ and for $\theta = \vec{a}$ are respectively*

$$(16)\qquad I_Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad I_Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1-\xi)^{-2} \end{pmatrix}.$$

Now the approach just sketched for the one-parameter case breaks down. Certainly we can form a preliminary estimator of θ and thereby 'estimate' the quantum score matrices. Next, in the full pure- and mixed-state models, one can rotate the coordinate system and reparametrize so that the quantum scores become σ_x, σ_y (pure-state model), and $\sigma_x, \sigma_y, a + b\sigma_z$ (mixed state model). There is no way, in either model, we can simultaneously measure these observables since they do not commute. Thus no measurement on a single particle has an information matrix equal to I_Q. The big question is, what *is* the class of information matrices I_M which are available? And if we can perform measurements on the system obtained by combining particles, what scaled information matrices $I_M^{(n)}/n$ become available? The latter class includes all of the former class, since the joint measurements include n i.i.d. copies of measurements on separate particles; moreover, these classes are convex and bounded.

Though all scaled information matrices $I_M^{(n)}/n$ are bounded by I_Q, we cannot expect them, for given n, to contain a single 'best' information. Which measurement we should choose, will depend on the relative accuracy with which we want to estimate the different components of θ. For instance, if in the pure-state case, close to $\theta = (0,0)$, we are only interested in θ_1, we should simply measure σ_x on each of our n particles yielding the maximum information on θ_1 but no information at all on θ_2. After we have characterized the class of all information matrices available, we must specify through some loss function the relative importance of the different parameters and solve some optimization problem.

2 A new Cramér-Rao type bound

In this section we report on recent results of Gill and Massar (2000), concentrating on the spin-half situation, and within that case, emphasizing the full pure-state model and the full mixed-state model. There turns out to be a striking difference between these two cases. For pure states, there is asymptotically no advantage in joint measurements on many particles. However for mixed states there typically is an advantage. How much is still an open question. The following result should be called a 'Theorem' (in quotes) since we do not specify regularity conditions and indeed only a 'Proof' exists, not yet a Proof.

Theorem 2.1 (Achievable information matrices, $n = 1$) *The set of all Fisher information matrices of outcomes of measurements of one spin-half particle for a smooth model $\rho(\theta)$ is $\{F : \mathrm{trace}(I_Q^{-1}F) \leq 1\}$.*

The parameter θ could be one-, two- or three-dimensional. We suppose either that we have a pure-state model, or a strictly mixed-state model. The argument, in Gill and Massar (2000), has two main parts. In the first part we show that for all M, $F = I_M$ satisfies trace$(I_Q^{-1}F) \leq d-1$ (we do not yet need that $d=2$). In the second part we show that, when $d=2$, for any F satisfying this inequality one can construct a measurement M for which $I_M = F$. For $d>2$, not all F satisfying trace$(I_Q^{-1}F) \leq 1$ are achievable, and it remains open to characterize exactly the class of achievable information matrices.

For the first part a series of preparatory steps are taken to bring us, 'without loss of generality', to a situation that allows exact computations. For simplicity take $d=2$. If $\rho(\theta)$ lies in the interior of the unit ball, and θ has dimension one or two, one can augment θ with other parameters, raising its dimension to 3. This can be done in such a way that the cross-information elements in the augmented $I_Q(\theta)$ are all zero. It then suffices to prove the inequality for θ of dimension 3, and then we may as well use the natural parametrization $\rho(\theta) = \frac{1}{2}(1+\vec{\theta}\cdot\vec{\sigma})$ with $\|\theta\| < 1$ since the quantity trace$(I_Q^{-1}F)$ is invariant under smooth reparametrization. If, on the other hand, $\rho(\theta)$ is a pure state model we can in the same way after augmenting θ assume that θ has dimension 2 and after reparametrization the model is $\rho(\theta) = \frac{1}{2}(1+\vec{\theta}\cdot\vec{\sigma})$ with $\|\theta\| = 1$.

For the next preparatory step we need the concepts of refinement and coarsening of a measurement.

Definition 2.1 (Coarsening and refinement) *A measurement M with sample space $\mathcal{X}$ is a refinement of M' with sample space $\mathcal{Y}$, and M' is a coarsening of M, if a measurable function $f : \mathcal{X} \to \mathcal{Y}$ exists with $M'(B) = M(f^{-1}(B))$.*

The result of measurement of M' then has the same distribution as taking f of the outcome of measurement of M. It follows that the Fisher information in the outcome of M' is less than or equal to that in M since under coarsening of data, Fisher information can only decrease.

Now we show that any measurement M' has a refinement M for which $M(A) = \int_A M(x)\mu(\mathrm{d}x)$ for some nonnegative operator-valued function M and bounded measure μ *and for which $M(x)$ has rank one for all x*, thus $M(x) = |x\rangle\langle x|$ for some (not necessarily normalised) vector function $|x\rangle$. Consequently it will suffice to prove the result for such *maximally refined* measurements M. Start with the measurement M' with sample space $\mathcal{Y}$. Define a probability measure ν on $\mathcal{Y}$ by $\nu(B) = \mathrm{trace}(M'(B))/d$; by taking Radon-Nikodym derivatives one can define $M'(y)$ such that $M'(B) = \int_B M'(y)\nu(\mathrm{d}y)$. Since the rank of $M'(y)$ is finite, $M'(y) = \sum_i M_i(y)$ where each $M_i(y)$ has rank one. Now refine the original sample space $\mathcal{Y}$ to $\mathcal{X} = \mathcal{Y} \times \{1,\ldots,d\}$, defining $M(A\times\{i\}) = \int_A M_i(y)\nu(\mathrm{d}y)$. Equivalently $M(A) = \int_A M_i(x)\mu(\mathrm{d}x,\mathrm{d}i)$ where μ is the product of ν with counting measure.

This brings us to the situation where the model is either full pure-state or full mixed-state, and where the measurement is maximally refined. We take the natural parametrization of either of these models, and without loss of generality work at a point θ where $\theta = (0,0)$ or $(0,0,\xi)$. This is possible by the result of Exercise 1.5. Now we have a formula for I_Q and for the derivatives of ρ with respect to the components of θ, both in the pure and the mixed case, and we have a representation for M in terms of a collection of vectors $|x\rangle$ which must satisfy the normalization constraint $\int_{\mathcal{X}} |x\rangle \langle x| \mu(\mathrm{d}x) = 1$ but which are otherwise arbitrary. Both ρ and I_Q are diagonal. We simply compute trace $I_Q^{-1} I_M$ and show that it equals 1 in the case $d = 2$. We leave the details as an exercise for the diligent reader—the computation is not difficult but does not seem all that illuminating either. We would dearly like to know if there is a more insightful way to get this result!

The same arguments work for arbitrary d though the details are more complicated; a full mixed-state model has $\frac{1}{2}d(d+1)$ parameters, a full pure-state model $\frac{1}{2}d(d+1) - (d-1)$ parameters, and a careful parametrization is needed to make I_Q diagonal.

In the second part (for $d = 2$ only) it is shown that for any F satisfying $\mathrm{trace}(I_Q^{-1}F) \leq 1$, one can construct a measurement M for which $I_M = F$. This measurement will be described in the next section. It typically depends on the point θ so a multi-stage procedure is going to be necessary to achieve asymptotically this information bound. That will be the main content of the next section, where we do some quantum asymptotics proving asymptotic optimality results for $n \to \infty$ of the resulting two-stage procedure.

We only have partial results for $n > 1$. In two special cases the available scaled information matrices do not increase as n increases. One of these cases is the case of pure-state models. This case has been much studied in the literature and is of great practical importance. The other case is when we make a restriction on the class of measurements to measurements of product form (in the literature also sometimes called an *unentangled measurement*). We first define this notion and then explain its significance.

Definition 2.2 (Product-form measurements) *We say that a measurement on n copies of a given quantum system is of product form if $M^{(n)}(A) = \int_A M^{(n)}(x)\mu(\mathrm{d}x)$ for a real measure μ and matrix-valued function $M^{(n)}(x)$ where $M^{(n)}(x)$ is of the form $M_1(x) \otimes \cdots \otimes M_n(x)$, with nonnegative components.*

We described in the previous section a measurement procedure whereby we first carried out measurements on some of our n particles, and then, depending on the outcome, carried out other measurements on the remaining particles. Altogether this procedure constitutes one measurement on the joint system of n particles taking values in some n-fold product space. One can conceive of more elaborate schemes where, depending on the results at any stage, one decides, possibly with the help of some classical randomisation, which particle to measure next and how. It would be allowed to measure again a particle which had previously been subject to measurement. There exists a general description of the state of a quantum system after measurement, allowing one to piece all the ingredients together into one measurement of the combined system. A measurement which can be decomposed into separate steps consisting of measurements on separate particles only, is called a *separable* measurement.

It turns out that all separable measurements, provided all outcomes of the component steps are encoded in the overall outcome x, have product-form. On the other hand, product-form measurements exist which are not separable; see Bennett et al. (1998). The product-form measurements form a large and interesting class, including all measurements which can be carried out sequentially on separate particles as well as more besides.

In the notion of separable measurement it is insisted that all intermediate outcomes are included in the final outcome. If one throws away some of the data, one gets an outcome whose distribution is the same as the distribution of a coarsening of the original measurement. Coarsening of a measurement can obviously destroy the product form, replacing products of nonnegative components by integrals of such products.

Theorem 2.2 (Achievable information matrices $n > 1$) *The scaled information matrices of measurements on a smooth model $\rho^{\otimes n}(\theta)$ remain $\{F : \mathrm{trace}(I_Q^{-1}F) \leq 1\}$*

1. *when θ is one-dimensional;*
2. *in a pure-state spin-half model;*
3. *in a mixed-state spin-half model with the class of measurements restricted to measurements which can be refined to product-form.*

The theorem is proved exactly as before, again finishing in an unilluminating calculation. Since coarsening of a measurement can only decrease Fisher information, one need only consider product-form measurements in the mixed-state case.

We have a counterexample to the conjecture that, for mixed states, the bound holds for *all* measurements. In the case $n = 2$, at the point $\rho = \frac{1}{2}\mathbf{1}$, there is a measurement for which $\mathrm{trace}(I_Q^{-1}I_M^{(2)}/2) = 3/2$, thus 50% more information in an appropriate measurement of two identical particles than any combination of separate measurements of the two. What the set of achievable scaled information matrices looks like and whether it continues to grow (and to what limit) as n grows, is completely unknown.

The measurement has seven elements, the first six of the form $\frac{1}{2}\Pi_{[\psi]}$, and the seventh $\Pi_{[\psi]}$, where $\Pi_{[\psi]}$ denotes the projector onto the one-dimensional sub-space spanned by the vector $|\psi\rangle$. The various $|\psi\rangle$ are $|{+z}+z\rangle$, $|{-z}-z\rangle$, $|{+x}+x\rangle$, $|{-x}-x\rangle$, $|{+y}+y\rangle$, $|{-y}-y\rangle$, $|S\rangle$. By $|{+z}+z\rangle$ we mean $|{+z}\rangle \otimes |{+z}\rangle = |\vec{e}_z\rangle \otimes |\vec{e}_z\rangle$ and similarly for the next five. The last ψ is the so-called *singlet state* $\frac{1}{\sqrt{2}}(|{+z}\rangle \otimes |{-z}\rangle - |{-z}\rangle \otimes |{+z}\rangle)$. As a pure state of two interacting spin-half particles, this is the famous entangled state resulting in the violation of the Bell inequalities, and hence of locality (according to some interpretations). Here it arises as part of a measurement of two completely non-interacting particles; however this measurement can never be implemented by doing separate operations on the separate particles.

Similar examples occur in the paper of Vidal et al. (1998), extending the pure-state results of Massar and Popescu (1995) to mixed states.

3 Quantum asymptotics

The results of the previous section are in the form of a bound on the information matrix based on the outcome of any measurement (perhaps restricted to the class of product-form measurements) on n identical copies of a given spin-half quantum system with state depending on an unknown parameter θ. We will now explain how such a bound can be used to give asymptotic bounds on the quality of estimators based on those measurements. Furthermore, we show how the bounds can be achieved by a two-stage procedure based on simple measurements on separate particles only. As far as achieving the bounds is concerned, only for the full mixed-state model under the natural parametrization is the problem completely solved. For the other models, the results are conjectural.

We will discuss two kinds of bounds: firstly, a bound on the limiting scaled mean quadratic error matrix of a well-behaved sequence of estimators, and secondly, a bound on the mean quadratic error matrix of the limiting distribution of a well-behaved sequence of estimators. Each has its advantages and disadvantages. In particular, since the delta-method works for (the variance of) limiting distributions but not for limiting mean square errors, stronger conditions are needed to prove optimality of some procedure in the first sense than in the second sense.

3.1 Two asymptotic bounds

Obviously a bound on the information matrix, by the ordinary Cramér-Rao inequality, immediately implies a bound on the covariance matrix of an unbiased estimator. However, this is not a restriction we want to make. It turns out much more convenient to work via a Bayesian version of the

Cramér-Rao inequality due to van Trees (1968), as generalised to the multiparameter case by Gill and Levit (1995). For a one-dimensional parameter the van Trees inequality is easy to state: the Bayes quadratic risk is bounded by one over expected information plus information in the prior. In the multiparameter case one has a whole collection of inequalities corresponding to different choices of quadratic loss function and some other parameters, more difficult to interpret.

Let $\pi(\theta)$ be a prior density for the p-dimensional parameter θ, which we suppose to be sufficiently smooth and supported by a compact and smoothly bounded region of the parameter space; see Gill and Levit (1995) for the precise requirements. Let $C(\theta)$ be a $p\times p$ symmetric positive-definite matrix (C stands for cost function) and let $V_M^{(n)}(\theta)$ be the mean quadratic error matrix of a chosen estimator of θ based on a measurement of n copies of the quantum system. Letting Θ denote a random drawing from the prior distribution π, it follows that $\mathrm{E}\,\mathrm{trace}\,C(\Theta)V_M^{(n)}(\Theta)$ is the Bayes risk of the estimator with respect to the loss function $(\widehat{\theta}^{(n)}-\theta)^\top C(\theta)(\widehat{\theta}^{(n)}-\theta)$.

Let $D(\theta)$ be another $p\times p$ matrix function of θ. Let $I_M^{(n)}(\theta)$ denote the Fisher information matrix in the measurement. Then the multivariate van Trees inequality reads

(17)

$$\mathrm{E}\,\mathrm{trace}\,C(\Theta)nV_M^{(n)}(\Theta) \geq \frac{(\mathrm{E}\,\mathrm{trace}\,D(\Theta))^2}{\mathrm{E}\,\mathrm{trace}\,C(\Theta)^{-1}D(\Theta)(I_M^{(n)}(\Theta)/n)D(\Theta)^\top + \widetilde{\mathcal{I}}(\pi)/n}$$

where

$$\widetilde{\mathcal{I}}(\pi) = \int \frac{1}{\pi(\theta)} \sum_{ijkl} C_{ij}^{-1}(\theta)\frac{\partial}{\partial\theta_k}(D_{ik}(\theta)\pi(\theta))\frac{\partial}{\partial\theta_l}(D_{jl}(\theta)\pi(\theta))\mathrm{d}\theta. \tag{18}$$

From Theorem 2.2 we have the bound $\mathrm{trace}\,I_Q^{-1}(\theta)(I_M^{(n)}(\theta)/n)\leq 1$, where, in the mixed case, we restrict attention to measurements refinable to product-form. We are going to assume that our sequence of measurements and estimators is such that the normalized mean quadratic error matrix $V_M^{(n)}(\theta)$ converges sufficiently regularly to a limit $V(\theta)$. Our aim is to transfer the inequality of Theorem (2.2) to V obtaining the bound $\mathrm{trace}\,I_Q^{-1}(\theta)V(\theta)^{-1}\leq 1$.

We will do this by making appropriate choices of C and D. We will need regularity conditions both on the sequence of estimators and on the model $\rho(\theta)$ in order to carry over (17) to the limit.

Theorem 3.1 (Asymptotic Cramér-Rao 1) *Suppose that on some open set of parameter values θ:*

1. *$nV_{(n)}$ converges uniformly to a continuous limit V;*

2. *$I_Q(\theta)$ is continuous with bounded partial derivatives;*

3. *V and I_Q are non-singular.*

Then the limiting normalised mean quadratic error matrix satisfies

$$\text{trace}\, I_Q^{-1}(\theta)V(\theta)^{-1} \leq 1. \tag{19}$$

We outline the proof of the theorem as follows. First of all, we pick a point θ_0 and define $V_0 = V(\theta_0)$. Next we define

$$C(\theta) = V_0^{-1}I_Q^{-1}(\theta)V_0^{-1}, \tag{20}$$

$$D(\theta) = V_0^{-1}I_Q^{-1}(\theta). \tag{21}$$

With these choices (17) becomes

(22)

$$\text{E}\,\text{trace}\, V_0^{-1}I_Q^{-1}(\Theta)V_0^{-1}(nV_M^{(n)}(\Theta)) \geq \frac{(\text{E}\,\text{trace}\, V_0^{-1}I_Q^{-1}(\Theta))^2}{\text{E}\,\text{trace}\, I_Q(\Theta)^{-1}(I_M^{(n)}(\Theta)/n) + \widetilde{\mathcal{I}}(\pi)/n}.$$

We can bound the first term in the denominator of the right hand side by 1, by the results of the last section. The second term in the denominator of the right hand side is finite, by our third assumption, and for $n \to \infty$ it converges to zero. By our first assumption (22) converges to

$$\text{E}\,\text{trace}\, V_0^{-1}I_Q^{-1}(\Theta)V_0^{-1}V(\Theta) \geq (\text{E}\,\text{trace}\, V_0^{-1}I_Q^{-1}(\Theta))^2. \tag{23}$$

Now replace the prior density π by one in a sequence of priors, concentrating on smaller and smaller neighbourhoods of θ_0. Using the continuity assumptions on V and I_Q, we obtain from (23) the inequality

$$\text{trace}\, V_0^{-1}I_Q^{-1}(\theta_0)V_0^{-1}V_0 \geq (\text{trace}\, V_0^{-1}I_Q^{-1}(\theta))^2.$$

or in other words, with $\theta = \theta_0$, the required

$$\text{trace}\, I_Q^{-1}(\theta)V^{-1}(\theta) \leq 1. \tag{24}$$

In some situations it might be more convenient to have a bound on the mean quadratic error of a limiting distribution, assuming one to exist. At the moment of writing we believe the following:

Theorem 3.2 (Asymptotic Cramér-Rao 2) *Let Z be distributed with the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta)$. Suppose that:*

1. *$\widehat{\theta}_n$ is Hájek regular at θ at root n rate;*

2. *the asymptotic mean quadratic error matrix* $V = \mathrm{E}(Z\,Z^{\top})$ *is non-singular;*

3. I_Q *is non-singular.*

Then V *satisfies*

$$\operatorname{trace} I_Q^{-1}(\theta)V(\theta)^{-1} \leq 1. \tag{25}$$

The proof should follow the lines of the similar result in Gill and Levit (1995), with a prior distribution concentrating on a root n neighbourhood of the truth. We will need similar choices of C and D as in the proof of Theorem 3.1, though the dependence of D on θ can now be suppressed.

3.2 Achieving the asymptotic bounds

At present we have essentially complete results in the full mixed-state spin-half model with the natural parametrization. We believe they can be extended to smooth (C^1) pure- and mixed-state models.

Give yourself a target mean quadratic error matrix $W(\theta)$ satisfying

$$\operatorname{trace} I_Q(\theta)^{-1}W(\theta)^{-1} \leq 1. \tag{26}$$

Is there a sequence of measurements $M^{(n)}$ satisfying the conditions of Theorems 3.1 or 3.2 with limiting mean quadratic error matrix $V(\theta)$ equal to the target?

Possibly we do not start with a target W but with a step earlier, with a quadratic cost function. For given $C(\theta)$ it is straightforward to compute the matrix $W(\theta)$ which minimizes $\operatorname{trace} C(\theta)W(\theta)$ subject to the constraint (26); the solution is $W = \operatorname{trace}((I_Q^{-\frac{1}{2}}CI_Q^{-\frac{1}{2}})^{\frac{1}{2}})I_Q^{-\frac{1}{2}}(I_Q^{\frac{1}{2}}CI_Q^{\frac{1}{2}})^{\frac{1}{2}}I_Q^{-\frac{1}{2}}$. Now we pose the same question again, with the W we have just calculated as target.

Let us call $F = W^{-1}$ the target information matrix. First we pretend θ is known and exhibit a measurement M on a single particle with the target information matrix at the given parameter value.

In the previous section we omitted explaining how the bound of Theorem 2.1 can be attained. That theorem stated that, at a given parameter value, for any positive-semidefinite symmetric F satisfying $\operatorname{trace} I_Q^{-1}F \leq 1$ there is a measurement M on a single spin-half particle with $I_M = F$. What is that measurement? We describe it in the case of a full mixed-state spin-half model with the natural parametrization, thus $\rho(\theta) = \frac{1}{2}(1 + \vec{\theta}\cdot\vec{\sigma})$. The matrices I_Q and F are 3×3.

To start with, we compute the eigenvector-eigenvalue decomposition of $I_Q^{-\frac{1}{2}}FI_Q^{-\frac{1}{2}}$, obtaining eigenvectors $\vec{h}_i$ and nonnegative eigenvalues γ_i, say. The condition on F translates to $\sum \gamma_i \leq 1$. Now define $\vec{g}_i = I_Q^{\frac{1}{2}}\vec{h}_i$ and three unit vectors $\vec{u}_i = g_i/\|g_i\|$, and finally consider the measurement M taking seven different values, whose elements are $\gamma_i \Pi(\pm\vec{u}_i)$, $i = 1,2,3$, and $(1 - \sum \gamma_i)\mathbf{1}$.

It turns out by a straightforward computation (carried out, without loss of generality, at $\theta = (0,0,\xi)$) that the information matrix for the measurement with the two elements $\Pi(\pm\vec{u}_i)$ has information matrix $\vec{g}_i \otimes \vec{g}_i$ and hence the measurement M has information matrix $\sum_i \gamma_i \vec{g}_i \otimes \vec{g}_i = F$.

This seven-outcome measurement can be implemented as a randomized choice between three simple measurements: with probability γ_i measure spin in the direction $\vec{u}_i$, with probability $1 - \sum \gamma_i$ do nothing.

However, in practice this measurement is not available, since the directions $\vec{u}_i$ and probabilities γ_i depend on the unknown θ. We therefore take recourse to the following two-stage measurement procedure.

First measure spin in the x, y and z directions on $\frac{1}{3}n^a$ each of the particles, where $0 < a < 1$ is fixed and the numbers are rounded to whole numbers. The expected relative frequency of 'up' particles in each direction is $\frac{1}{2}(1 + \theta_i)$, $i = 1,2,3$, so solving 'observed equals expected' yields a consistent preliminary estimator $\widetilde{\theta}$ of θ. If the estimate lies outside the unit-ball, project onto the ball, and stop. With large probability no projection is necessary. We can compute the eigenvalue-eigenvector decomposition of $I_Q^{-\frac{1}{2}}(\widetilde{\theta})F(\widetilde{\theta})I_Q^{-\frac{1}{2}}(\widetilde{\theta})$, leading to fractions γ_i and directions $\vec{u}_i$ as above. Measure the spin of a fraction γ_i of the remaining particles in the direction $\vec{u}_i$. Solve again the three (linear) equations 'observed relative frequency equals expected', treating the $\vec{u}_i$ as fixed. Project onto the unit ball if necessary. Call the resulting estimator $\widehat{\theta}$.

Our claim is that this procedure exhibits a measurement $M^{(n)}$ on the n particles, and an estimator $\widehat{\theta}^{(n)}$ based on its outcome, which satisfies the conditions of Theorem 3.1, with $V(\theta)$ equal to the target $W(\theta)$. Thus the bound of Theorem 3.1 is also achievable, and a measurement which does this has been explicitly described above. Apart from projecting onto the unit ball, the estimator involves only linear operations on binomial variables, so is not difficult to analyse explicitly. We need a preliminary sample size $\widetilde{n}$ of order n^a and not, for example, of order $\log n$, in order to control the scaled mean quadratic error of the estimator. There is an exponentially small probability—in $\widetilde{n}$, not in n—that the preliminary estimate is outside of a given neighbourhood of the truth, and hence that the scaled quadratic error is of order n.

One can further check that the estimator we have described also satisfies the conditions of Theorem 3.2.

Possibly one is interested in a different parametrization of the model. Under a smooth (C^1) reparametrization, the delta method allows us to maintain optimality in the sense of Theorem 3.2. However optimality in the sense of Theorem 3.1 could be destroyed; in order for it to be maintained the reparametrization should also be bounded. Alternatively one must modify the estimator by a truncation at a level increasing slowly enough to infinity with n, cf. Schipper (1997; section 4.4) or Levit and Oudshoorn (1993) for examples of the technique.

This approach can be extended to other spin-half models. The difficulties are exemplified by the case of the two-parameter full pure-state spin-half model. Locally, consider the natural parametrization $\theta = (\theta_1, \theta_2)$, $\theta_3 = (1 - \theta_1^2 - \theta_2^2)^{1/2}$, $\rho = \rho(\vec{\theta})$ at the point $\theta = (0, 0)$. The quantum information matrix for three parameters $\theta_1, \theta_2, \theta_3$ contains an infinite element. However, the recipe outlined above continues to work if we add to a given 2×2 target information matrix a third zero row and column—infinities always get multiplied by zero. The third fraction γ_3 equals zero so simple measurements in just two directions suffice.

The resulting procedure involves linear operations on binomial counts, projecting onto S, and reparametrization. Under some smoothness we should finish with an estimator optimal in the sense of Theorem 3.2; under further smoothness, boundedness, and a sufficiently large preliminary sample also optimality in the sense of Theorem 3.1 should hold.

If the target information matrix includes some zeros, i.e., one is not interested at all in certain parameters, the results should still go through; the preliminary sample should be of size of order n^a, $\frac{1}{2} < a < 1$, in order that the uncertainty in the initial estimate of the 'nuisance parameters' does not contaminate the final result.

4 Non-locality without entanglement

It would take us too far afield here to explain the notions of entanglement and of non-locality. For some kind of introduction see Kümmerer and Maassen (1998) and Gill (1998), and Gill (2000); see also the books of Peres (1995), Isham (1995), Penrose (1994), Maudlin (1994). However, we would like to discuss whether or not our finding that non-separable joint measurements on several independent (non-entangled) quantum particles can yield more information than any separate measurements on the separate particles, should be considered surprising or not. Recall that separable measurements, cf. Bennett et al. (1998), are measurements which can be decomposed into a sequence of measurements on separate particles, each measurement possibly depending on the outcome of the preceding ones, and whereby it is allowed to measure further a particle which has already been measured (and hence its state has been altered in a particular way) at an earlier step.

From a mathematical point of view there should not be much surprise. The class of separable measurements is contained in the class of product-form measurements, which is clearly a very small part of the space of all measurements whatsoever. The optimisation problem of maximising Fisher information (more precisely, some scalar functional thereof) must only be expected to have a larger outcome when we optimise over a larger space. The surprise for the mathematician is rather that for pure states, and for one dimensional parameters, there is no gain in joint measurements. And it is strange that mixed states should exhibit this phenomenon, whereas pure states do not: the difference is classical probabilistic mixing, which should not lead to nonclassical behaviour.

However, physicists are and should be surprised. The reason is connected to the feeling of many physicists that the randomness in measurement of a quantum system should have a deterministic explanation (Einstein: "God does not throw dice") . We appreciate very well that tossing a coin is essentially a completely deterministic process. It is only uncontrolled variability in initial conditions which lead to the outcome appearing to be completely random. Might it be the case also that the randomness in the outcome of a measurement of a quantum system might be 'merely' the reflection of statistical variability in some initial conditions? Hidden variables, so called because at present no physicist is aware what these lower level variables are, and there is no known way directly to measure them.

In fact there already exist arguments aplenty that if there *is* a deterministic hidden layer beneath quantum theory, it violates other cherished physical intuitions, in particular the principle of locality; see again Kümmerer and Maassen (1998), Gill (1998) for some introduction to the phenomenon of entanglement, and further references. But let us ignore that evidence and consider the new evidence from the present results. Consider two identical copies of a given quantum state. Suppose there were a hidden deterministic explanation for the randomness in the outcome of any measurement on either or both of these particles. Such an explanation would involve hidden variables ω_1, ω_2 specifying the hidden state of the two particles. Since applying separate measurements to the two systems produces independent outcomes, and since the outcomes of the same measurements are identically distributed, one would naturally suppose that these two variables are independent and identically distributed. Their distributions would of course depend on the unknown parameter θ. Now when we measure the joint system, there could be other sources of randomness in our experiment, possibly even quantum randomness, but still it would not have a distribution depending on θ. So let us assume there is a third random element ω_M such that the outcome of the measurement M on the system $\rho(\theta) \otimes \rho(\theta)$ is a deterministic function of ω_1, ω_2 and ω_M; the first two are independent and identically distributed, with marginal distributions depending on θ, while the distribution of ω_M given the other two is independent of θ. Thus the random outcome X of the measurement of M is just $X(\omega_1, \omega_2, \omega_M)$, a random variable on the probability space $(\Omega \times \Omega \times \Omega_M), ((\mathrm{P}_\theta \times \mathrm{P}_\theta) * \mathrm{P}_M)$, where P_M is some Markov kernel from $\Omega \times \Omega$ to Ω_M. Now it is well-known from ordinary statistics that the Fisher information in θ from the distribution of any random variable defined on this space is less than twice the information in one observation of ω_1 itself seen as a random variable defined on $(\Omega, \mathrm{P}_\theta)$. Thus if one could realise any Ω_M, P_M and any X whatsover by suitable choice of measurement M, achievable Fisher information would be additive!

What can we conclude from the fact that achievable Fisher information is not additive? If they exist, the hidden variables are so well hidden that we cannot uncover them from any measurements on single particles, i.e., it is not possible to realise any (Ω_M, P_M) and any X whatever by appropriate choice of experimental set-up. However, we can uncover the hidden variables better, apparently, from appropriate measurements on several particles brought together, even though these particles have nothing whatever to do with one another—their hidden variables are independent and identically distributed. Alternatively the explanation must be found in some pathological non-measurability or non-regularity of the statistical model we have just introduced. Whatever escape-route one chooses, it is clear that *if* there is a deterministic explanation for quantum randomness, it has to be a very weird explanation. God throws rather peculiar dice.

Acknowledgements. This paper is based on work in progress together with O.E. Barndorff-Nielsen and with S. Massar. I am grateful for the hospitality of the Department of Mathematics and Statistics, University of Western Australia. I would like to thank Boris Levit for his patient advice.

REFERENCES

Barndorff-Nielsen, O.E. and Gill, R.D. (2000). Fisher information in quantum statistics. To appear in *J. Phys. A*. Preprint `quant-ph/9802063, http://xxx.lanl.gov`

Belavkin, V. P. (1976). Generalized uncertainty relations and efficient measurements in quantum systems. *Teoreticheskaya i Matematiskaya Fizika* **26**, 316–329; English translation, 213–222. Based on author's candidate's dissertation, Moscow State University, 1972.

Bennett, C.H., DiVincenzo, D.P., Fuchs, C.A., Mor, T., Rains, E., Shor, P.W., Smolin, J.A., and Wootters, W.K. (1998). Quantum nonlocality without entanglement. To appear in *Phys. Rev. A*. Preprint `quant-ph/9804053, http://xxx.lanl.gov`

Biane, P. (1995). Calcul stochastique non-commutatief. pp. 4–96 in: *Lectures on Probability Theory: Ecole dét'e de Saint Flour XXIII–1993*, P. Biane and R. Durrett, Springer Lecture Notes in Mathematics **1608**.

Braunstein, S.L. and Caves, C.M. (1994). Statistical distance and the geometry of quantum states. *Physical Review Letters* **72**, 3439–3443.

Brody, D.C. and Hughston, L.P. (1998), Statistical geometry in quantum mechanics, *Proceedings of the Royal Society of London Series A* **454**, 2445–2475.

Gill, R.D. (2000). *Lecture Notes on Quantum Statistics.* `http://www.math.uu.nl/people/gill/Preprints/book.ps.gz`

Gill, R.D. (1998). Critique of 'Elements of quantum probability'. *Quantum Probability Communications* **10**, 351–361.

Gill, R.D. and Levit, B.Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli* **1** 59–79.

Gill, R.D. and Massar, S. (2000). State estimation for large ensembles. To appear in *Phys. Rev. A.* Preprint `quant-ph/9902063, http://xxx.lanl.gov`

Helstrom, C.W. (1967). Minimum mean-square error of estimates in quantum statistics. *Phys. Letters* **25 A**, 101–102.

Helstrom, C.W. (1976). *Quantum Detection and Estimation Theory.* Academic, New York.

Holevo, A.S. (1982). *Probabilistic and Statistical Aspects of Quantum Theory.* North Holland, Amsterdam.

Holevo, A.S. (1983). Bounds for generalized uncertainty of the shift parameter. *Springer Lecture Notes in Mathematics* **1021**, 243–251.

Holevo, A.S. (1999). *Lectures on Statistical Structure of Quantum Theory.* `http://www.imaph.tu-bs.de/skripte_.html`

Isham, C. (1995). *Quantum Theory.* World Scientific, Singapore.

Kümmerer, B. and Maassen, H. (1998). Elements of quantum probability. *Quantum Probability Communications* **10**, 73–100.

Levit, B.Y. and Oudshoorn, C.G.M. (1993). Second order admissible estimation of variance. *Statistics and Decisions, supplement issue* **3**, 17–29.

Malley, J.D. and Hornstein, J. (1993). Quantum statistical inference. *Statistical Science* **8**, 433–457.

Massar, S. and Popescu, S. (1995). Optimal extraction of information from finite quantum ensembles. *Physical Review Letters* **74**, 1259–1263.

Maudlin, T. (1994). *Quantum Non-locality and Relativity.* Blackwell, Oxford.

Meyer, P.A. (1986). Eléments de probabilités quantiques. pp. 186–312 in: *S'eminaire de Probabilités XX*, ed. J. Azéma and M. Yor, *Springer Lecture Notes in Mathematics* **1204**.

Penrose, R. (1994). *Shadows of the Mind: a Search for the Missing Science of Consciousness.* Oxford University Press.

Percival, I. (1998). *Quantum State Diffusion.* Cambridge University Press.

Peres, A. (1995). *Quantum Theory: Concepts and Methods.* Kluwer, Dordrecht.

Peres, A. and Wootters, W.K. (1991). Optimal detection of quantum information. *Physical Review Letters* **66**, 1119–1122.

Preskill (1997). *Quantum Information and Computation.*
`http://www.theory.caltech.edu/people/preskill/ph299`

Schipper, C.M.A. (1997). *Sharp Asymptotics in Nonparametric Estimation.* PhD thesis, University Utrecht, ISBN 90-393-1208-7.

Stratonovich, R.L. (1973). The quantum generalization of optimal statistical estimation and testing hypotheses. *Stochastics* **1**, 87–126.

van Trees, H.L. (1968). *Detection, Estimation and Modulation Theory (Part 1).* Wiley, New York.

Vidal, G., Latorre, J.I., Pascual, P., and Tarrach, R. (1998). *Optimal minimal measurements of mixed states.*
Preprint `quant-ph/9812068`, `http://xxx.lanl.gov`.

R.F. Werner (1997). *Quantum information and Quantum Computing.*
`http://www.imaph.tu-bs.de/skripte_.html`

MATHEMATICAL INSTITUTE
UNIVERSITY OF UTRECHT
BOX 80010
3508 TA UTRECHT
NETHERLANDS
gill@math.uu.nl