On Quantum Statistical Inference*

Ole E. Barndorff-Nielsen

MaPhySto†, University of Aarhus, Denmark.

Richard D. Gill

Mathematical Institute, University of Utrecht and EURANDOM, Eindhoven, Netherlands

Peter E. Jupp

School of Mathematics and Statistics, University of St Andrews, U.K.

Summary. Recent developments in the mathematical foundations of quantum mechanics have brought the theory closer to that of classical probability and statistics. On the other hand, the unique character of quantum physics sets many of the questions addressed apart from those met classically in stochastics. Furthermore, concurrent advances in experimental techniques and in the theory of quantum computation have led to a strong interest in questions of quantum information, in particular in the sense of the amount of information about unknown parameters in given observational data or accessible through various possible types of measurements. This scenery is outlined.

1. Introduction

In the last two decades, developments of an axiomatic type in the mathematical foundations of quantum mechanics have brought the theory closer to that of classical probability and statistics. On the other hand, the unique character of quantum physics (we use the terms 'quantum mechanics' and 'quantum physics' synonymously) sets many of the questions addressed apart from those met classically in stochastics. The key mathematical notion is that of a *quantum instrument*, which we shall describe in Section 2 and which, for arbitrary quantum experiments, specifies the joint probability distribution of the observational outcome of the experiment together with the state of the physical system after the experiment. Concurrently with these theoretical developments, major advances in experimental techniques have opened many possibilities for studying small quantum systems and this has led to considerable current interest in a range of questions that in essence belong to statistical inference and are concerned with the amount of information about unknown parameters in given observational data or accessible through various possible types of measurements. In quantum physics, the realm of possible experiments is specified mathematically, and noncommutativity between experiments plays a key role. Separate measurements on independent and separate systems result in independent observations, as in classical stochastics. However, joint measurements allow for major increases in statistical information.

The present paper outlines some of these developments and contains suggestions for additional readings and further work. We make some new contributions to the theory of quantum statistical inference, in particular, developing new notions of quantum sufficiency and exhaustivity. We give

*Version: 21/07/01

†MaPhySto is the Centre for Mathematical Physics and Stochastics, funded by the Danish National Research Foundation

complete but short proofs of the quantum information (Cramér–Rao) bound and some of its consequences, filling some gaps in the proofs in the physics literature. The paper does *not* contain practical examples in the sense of real data analyses, for several reasons. For one thing, the realistic modelling of present-day laboratory experiments in this field involves several more layers of complexity (technical, not conceptual) on top of the picture presented here. The closest we come to real data is in our discussion of quantum tomography in Section 7.2. For another thing, the theory in this paper is largely concerned with the design rather than the analysis of experiments in quantum physics, and there is still a gap between what is theoretically possible under the laws of quantum mechanics, and what is practically possible in the laboratory, though this gap is closing fast. 'Information' is understood throughout in the sense it has in mathematical statistics. We do not discuss quantum information theory in the sense of optimal coding and transmission of messages through quantum communication channels, nor in the more general sense of quantum information processing (Green, 2000). Within quantum statistics, we concentrate on the topics of estimation and of inference. The classic books of Helstrom (1976) and Holevo (1982) are on the other hand largely devoted to a decision theoretic approach to hypothesis testing problems. See Parthasarathy (1999) and Ogawa and Nagaoka (2000) for recent contributions to this field. Confusingly, the phrase 'maximum likelihood estimator' has an unorthodox meaning in the older literature. In many papers of which we just mention a few recent ones, Belavkin (1994, 2000, 2001) develops a continuous time Bayesian filtering approach to estimation and control.

It should be emphasised from the start, that we see quantum mechanics as describing *classical probability models* for the outcomes of laboratory experiments, or indeed, for the real world outcomes of any interactions between 'the quantum world' of microscopic particles and 'the real world' in which statisticians analyse data. Those probability models may depend on unknown parameters, and quantum statistics is concerned with statistical design and inference concerning those parameters. This point of view is commonplace in experimental quantum physics but seems to be less common in theoretical physics and in some parts of pure mathematics, in particular in the field called 'quantum probability', where a special nature is claimed for the randomness of quantum mechanics, placing it outside the ambit of classical probability and statistics. We disagree firmly with this conclusion though we do agree that there are fascinating foundational issues in quantum mechanics. We develop our stance on these issues further in Section 8. Quantum mechanics is concerned with randomness of the most fundamental nature known to science, and probabilists and statisticians definitely should be be involved in the game, rather than excluded from it.

In quantum mechanics the state of a physical system is described by a non-negative self-adjoint operator ρ (referred to as the *state*) with trace 1, on a separable complex Hilbert space $\mathcal{H}$. In accordance with the previous paragraphs, our interest in this paper concerns cases where the state is specified only up to some unknown parameter θ and the question is what can be learned about the parameter from observation of the system.

Many of the central ideas can be illustrated by finite-dimensional quantum systems, the simplest being based on those in which $\mathcal{H}$ has (complex) dimension 2. We shall often use the phrase 'spin-half particle' to refer to such a quantum system, as one of the best known examples concerns the magnetic moment or spin of the electron, which in appropriate units can only take on the values $\pm\frac{1}{2}$. But a two-dimensional state space is also appropriate for modelling the polarisation of one photon, and yet another example is provided by an atom at very low temperature when only its ground state and first excited state are relevant. The theory of quantum computation is concerned with how a finite collection of two-dimensional quantum systems, which are then called *qubits*, can be used to carry and manipulate information. We shall mainly concentrate on such examples. However, many physical problems concern infinite-dimensional systems, one area of great current interest being quantum tomography and quantum holography, which we shall discuss briefly. While the theory for finite-dimensional systems can be outlined in relatively simple mathematical terms, in general it is necessary to draw on advanced aspects of the theory of operators on infinite-dimensional Hilbert spaces and we will only outline this, with quantum tomography in mind, in Section 7.

The paper is organised as follows. Section 2 describes the mathematical structure linking states of a quantum system, possible measurements on that system, and the resulting state of the system after measurement. Section 3 introduces quantum statistical models and notions of quantum score and quantum information, parallel to the score function and Fisher information of classical parametric statistical models. In Section 4 we introduce quantum exponential models and quantum transformation models, again forming a parallel with fundamental classes of models in classical statistics. In Section 5 we describe the notions of quantum exhaustivity and quantum sufficiency of a measurement, relating them to the classical notion of sufficiency. We next, in Section 6, turn to a study of the relation between quantum information and classical Fisher information, in particular through Cramér–Rao type information bounds. In Section 7 we discuss the infinite-dimensional model of quantum tomography, which poses the challenge of developing non-parametric quantum information bounds. In Section 8 we discuss the difference between classical and quantum probability and statistics, relating them both to foundational issues in quantum physics and to emerging quantum technologies. Finally in Section 9 we conclude with remarks on further topics, in particular, quantum stochastic processes. The appendix contains some mathematical details.

This paper greatly extends our more mathematical survey (Barndorff-Nielsen, Gill, and Jupp, 2001) on quantum statistical information. Gill (2001a,b) contains further introductory material. Many proofs and further details will be found in Barndorff-Nielsen, Gill, and Jupp (2002). Some general references which we have found extremely useful are the books of Isham (1995), Peres (1995), Gilmore (1994), Holevo (1982, 2001c). Finally, the Los Alamos National Laboratory preprint service for quantum physics, `quant-ph` at `http://xxx.lanl.gov` is an invaluable resource.

2. States, Measurements and Instruments

In quantum mechanics the *state* of any physical system to be investigated is described by an operator ρ on a complex separable Hilbert space $\mathcal{H}$ such that ρ is non-negative and (hence) self-adjoint and has trace 1. In this paper (except for Section 7) we shall restrict attention to the case where $\mathcal{H}$ is finite-dimensional, and our examples will mainly concern the spin of spin-half particles, where the dimension of $\mathcal{H}$ is 2. The classic example in this context is the 1922 experiment of Stern and Gerlach, see Brandt and Dahmen (1995, Section 1.4), to determine the size of the magnetic moment of the electron. The electron was conceived of as spinning around an axis and therefore behaving as a magnet pointing in some direction. Mathematically, each electron carries a vector 'magnetic moment'. One might expect the size of the magnetic moment of all electrons to be the same, but the directions to be uniformly distributed in space. Stern and Gerlach made a beam of silver atoms move transversely through a steeply increasing vertical magnetic field. A silver atom has 47 electrons but it appears that the magnetic moments of the 46 inner electrons cancel and essentially only one electron determines the spin of the whole system. Classical physical reasoning predicts that the beam would emerge spread out vertically according to the component of the spin of each atom (or electron) in the direction of the gradient of the magnetic field. The spin itself would not be altered by passage through the magnet. However, amazingly, the emerging beam consisted of just two well separated components, as if the component of the spin vector in the vertical direction of each electron could take on only two different values.

In this case, $\mathcal{H}$ can be thought of as $\mathbb{C}^2$, i.e. as pairs of complex numbers, and, correspondingly,

ρ is a 2×2 matrix

$$\begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$$

with $\rho_{21} = \overline{\rho}_{12}$ (the bar denoting complex conjugation) and non-negative real eigenvalues p_1 and p_2 satisfying $p_1 + p_2 = 1$.

The result of performing a measurement on the system in state ρ is a random variable x taking values in a measure space $(\mathcal{X}, \mathcal{A})$ and with law of the form

$$\Pr(x \in A) = \operatorname{tr}\{\rho M(A)\}\,,$$

where M is a mapping from the σ-algebra $\mathcal{A}$ into the space $\mathbb{SA}_+(\mathcal{H})$ of non-negative self-adjoint operators on $\mathcal{H}$ which satisfies $M(\mathcal{X}) = \mathbf{1}$ (where $\mathbf{1}$ is the identity operator) and

$$\sum_{i=1}^{\infty} M(A_i) = M(A)$$

for any finite or countable sequence $\{A_1, A_2, \dots\}$ of disjoint elements of $\mathcal{A}$ and $A = \cup_{i=1}^{\infty} A_i$ (the sum in the formula being defined in the sense of weak convergence of operators). Such a mapping M is said to be a (generalised) *measurement*. We shall also refer to M as an *operator-valued probability measure* or *OProM* for short. In the literature the usual names and acronyms are probability operator-valued measure or positive operator-valued measure (POM or POVM), and (nonorthogonal, generalised) resolution of the identity.

The most basic measurements, which are among the class of *simple measurements* defined in Section 2.2, have $\mathcal{X}$ a finite set of real numbers, with cardinality less than or equal to the dimension of $\mathcal{H}$, $\mathcal{A}$ as the σ-algebra of all subsets of $\mathcal{X}$, $M(\{x\}) = \Pi_{[x]}$ for any atom $\{x\}$ of $\mathcal{A}$, the $\Pi_{[x]}$ being mutually orthogonal *projection operators* with $\sum \Pi_{[x]} = \mathbf{1}$. We speak then of a *projector-valued probability measure* or *PProM*. The usual terminology in the literature is a PVM or (orthogonal) resolution of the identity. All the ingredients of such a simple measurement are encapsulated in the specification of a self-adjoint operator Q on $\mathcal{H}$ with eigenvalues x in $\mathcal{X}$ and eigenspaces which are precisely those subspaces onto which the $\Pi_{[x]}$ project. The operator $Q = \sum x \Pi_{[x]}$ is called the *observable*. Conversely, any self-adjoint operator on $\mathcal{H}$ can be given an interpretation as an observable. We denote the space of self-adjoint operators (observables) by $\mathbb{SA}(\mathcal{H})$ and the set of states ρ by $\mathcal{S}(\mathcal{H})$. The adjoint of an operator is indicated by an asterisk $*$.

Physics textbooks on quantum theory usually take the concept of observables as a starting point. In the infinite dimensional case, observables—self-adjoint operators, not necessarily bounded—may have continuous spectrum instead of discrete eigenvalues. But the one-to-one correspondence between PProM's and observables continues to hold. Any self-adjoint operator on $\mathcal{H}$ can be given an interpretation as an observable.

Let M be a measurement. We shall often assume that M is dominated by a σ-finite measure ν on $(\mathcal{X}, \mathcal{A})$ and we shall write $m(x)$ for the density of M with respect to ν. Thus

$$M(A) = \int_A m(x)\nu(\mathrm{d}x)\,.$$

We can take $m(x)$ to be self-adjoint and nonnegative for all x. (If $\mathcal{H} = \mathbb{C}^d$, then $M(A)$ and $m(x)$ may be considered as $d \times d$ matrices of complex numbers.) The law of x is also dominated by ν and the probability density function of x is

$$p(x) = \operatorname{tr}\{\rho m(x)\}\,.$$

The physical state may depend on an unknown parameter θ, which runs through some parameter space Θ. In this case we denote the state by $\rho(\theta)$. Then the law of the outcome x of a measurement M depends on θ as well and we indicate this by writing $\mathrm{P}_\theta(A)$ or $p(x;\theta)$ for the probability or the probability density, as the case may be. In particular,

$$p(x;\theta) = \mathrm{tr}\{\rho(\theta)m(x)\}\,. \tag{1}$$

It may also be relevant to stress the dependence on M and we then write $p(x;\theta;M)$, etc. We shall refer to the present kind of setting as a *parametric quantum model* $(\boldsymbol{\rho}, M)$ or $(\boldsymbol{\rho}, m)$ with elements $\boldsymbol{\rho} = (\theta \mapsto \rho(\theta))$ and M, or its density m. It is also relevant to consider cases where the measurement M depends on some unknown parameter, but we shall not discuss this possibility further in the present paper. When the measurement M is given, a problem of classical statistical inference results concerning the model (1) for the distribution of the outcome. However, it turns out that the model for the state $\theta \mapsto \rho(\theta)$ can be usefully studied independently of which measurement is made of the system (or in order to choose the best measurement) and then quantum analogues of many concepts from classical statistical inference become important.

OProM's specify the probabilistic law of the outcome of an actual measurement but do not say anything about the state of the physical system after the measurement has been performed. The mathematical concept of *quantum instrument* prescribes both the OProM for the measurement and the posterior state.

The next three subsections discuss in more detail the concepts of states, measurements (or OProM's), and quantum instruments.

2.1. States

As stated at the beginning of the section, the state of a quantum system is represented by an *operator* ρ in $\mathcal{S}(\mathcal{H})$. It is often called the *density matrix* or *density operator* of the system. We think of vectors ψ in $\mathcal{H}$ as column vectors, and will emphasise this by writing $|\psi\rangle$ (Dirac's 'ket' notation). The adjoint (complex conjugate and transpose) of $|\psi\rangle$ is a row vector, which we denote by $\langle\psi|$ (Dirac's 'bra' notation).

The simplest states, called *pure states*, are the projectors of rank one, i.e. they are of the form $\rho = |\psi\rangle\langle\psi|$, where ψ is a unit vector in $\mathcal{H}$ (so $\langle\psi|\psi\rangle = 1$), called the *state-vector* of the pure state ρ. If $\mathcal{H}$ has dimension d then the set $\mathcal{S}_1(\mathcal{H})$ of pure states can be identified with the complex projective space $\mathbb{C}P^{d-1}$. In particular, $\mathcal{S}_1(\mathbb{C}^2)$ can be identified with the sphere S^2, which is known in theoretical physics as the *Poincaré sphere*, in quantum optics as the *Bloch sphere*, and in complex analysis as the *Riemann sphere*.

Example 1 (Spin-half) Take $\mathcal{H} = \mathbb{C}^2$, so that $\mathcal{H}$ has complex dimension 2, the space of general operators on $\mathcal{H}$ has real dimension 8, and the space $\mathbb{SA}(\mathcal{H})$ of self-adjoint operators on $\mathcal{H}$ has real dimension 4.

The space $\mathbb{SA}(\mathcal{H})$ is spanned by the identity matrix

$$\mathbf{1} = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

together with the *Pauli matrices*

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

Note that σ_x, σ_y and σ_z satisfy the *commutativity relations*

$$\begin{aligned}[\sigma_x, \sigma_y] &= 2i\sigma_z \\ [\sigma_y, \sigma_z] &= 2i\sigma_x \\ [\sigma_z, \sigma_x] &= 2i\sigma_y\end{aligned}$$

where, for any operators A and B, their commutator $[A, B]$ is defined as $AB - BA$; and note that

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbf{1}\,.$$

Any pure state has the form $|\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle$ in $\mathbb{C}^2$. Up to a complex factor of modulus 1 (the *phase*, which does not influence the state), we can write $|\psi\rangle$ as

$$|\psi\rangle = \begin{pmatrix} e^{-i\varphi/2}\cos(\vartheta/2) \\ e^{i\varphi/2}\sin(\vartheta/2) \end{pmatrix}.$$

The corresponding pure state is

$$\rho = \begin{pmatrix} \cos^2(\vartheta/2) & e^{-i\varphi}\cos(\vartheta/2)\sin(\vartheta/2) \\ e^{i\varphi}\cos(\vartheta/2)\sin(\vartheta/2) & \sin^2(\vartheta/2) \end{pmatrix}.$$

A little algebra shows that ρ can be written as $\rho = (\mathbf{1} + u_x\sigma_x + u_y\sigma_y + u_z\sigma_z)/2 = \frac{1}{2}(\mathbf{1} + \vec{u}\cdot\vec{\sigma})$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the three Pauli spin matrices and $\vec{u} = (u_x, u_y, u_z) = \vec{u}(\vartheta, \varphi)$ is the point on S^2 with polar coordinates (ϑ, φ). □

2.1.1. Mixing and Superposition

There are two important ways of constructing new states from old. Firstly, since the set of states is convex, new states can be obtained by mixing states $\rho_1, \ldots, \rho_m$, i.e. taking convex combinations

$$p_1\rho_1 + \cdots + p_m\rho_m\,, \tag{2}$$

where $p_1, \ldots, p_m$ are real with $p_i \geq 0$ and $p_1 + \cdots + p_m = 1$. If $\mathcal{H}$ is finite-dimensional then all states are of the form (2) with the ρ_i pure, so that $\mathcal{S}(\mathcal{H})$ is the convex hull of $\mathcal{S}_1(\mathcal{H})$: in the infinite-dimensional case one needs infinite mixtures. For this reason, states which are not pure are called *mixed states*. In particular, if $\mathcal{H} = \mathbb{C}^2$ then the set of pure states is the Poincaré sphere, whereas the set of mixed states is the interior of the corresponding unit ball.

If $\mathcal{H} = \mathbb{C}^d$ then mixing the pure states by the uniform probability measure on $\mathbb{C}P^{d-1}$ gives a state which is invariant under the action $\rho \mapsto U\rho U^*$ of SU(d), the group of special (determinant $+1$) unitary ($UU^* = U^*U = \mathbf{1}$) matrices of order d; this is the unique such invariant state.

The other important way of constructing new states from old is by superposition. The *superposition principle* states that a complex linear combination of state-vectors is also a physically possible state-vector. Let $|\psi_1\rangle\langle\psi_1|, \ldots, |\psi_m\rangle\langle\psi_m|$ be pure states on $\mathcal{H}$. Then any state which can be written in the form $\langle\psi|\psi\rangle^{-1}|\psi\rangle\langle\psi|$, where

$$\psi = w_1\psi_1 + \cdots + w_m\psi_m$$

and $w_1, \ldots, w_m$ are some complex numbers, is called a *superposition* of the pure states with state-vectors $|\psi_1\rangle, \ldots, |\psi_m\rangle$ (here the phases of the state-vectors are relevant!).

The difference between superposition and mixing may be illustrated by a spin-half example: take $\langle\psi_1| = (1, 0)$ and $\langle\psi_2| = (0, 1)$. For the superposition with $w_1 = w_2 = 1/\sqrt{2}$, we have

$$\langle\psi|\psi\rangle^{-1}|\psi\rangle\langle\psi| = \tfrac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix},$$

whereas the mixed state

$$p_1|\psi_1\rangle\langle\psi_1| + p_2\,|\psi_2\rangle\langle\psi_2| = \tfrac{1}{2}\begin{pmatrix}p_1 & 0\\ 0 & p_2\end{pmatrix}$$

is different from the preceding superposition, whatever p_1 and $p_2 = 1 - p_1$. Taking $p_1 = p_2 = \frac{1}{2}$, if we measure the PProM defined by the two projectors $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ and corresponding outcomes $+1$ and -1, the two states are indistinguishable: each gives probabilities of $1/2$ for the two outcomes. However, if we measure $\langle\psi|\psi\rangle^{-1}|\psi\rangle\langle\psi|$ and $|\psi\rangle^{\perp}\langle\psi|^{\perp}$, where $|\psi\rangle^{\perp}$ denotes a unit vector in $\mathbb{C}^2$ orthogonal to $|\psi\rangle$, then the second state again gives each outcome probability half, while the first state gives probabilities 1 and 0.

The possibility of taking complex superpositions of state-vectors to get new pure states corresponds to the wave-particle duality at the heart of quantum mechanics (linear combinations of solutions to wave equations are also solutions to wave equations). The new states obtained in this way will have distinctively different properties from the states out of which they are constructed. On the other hand, taking mixtures of states represents no more and no less than ordinary probabilistic mixtures: with probability p_i the system has been prepared in state ρ_i, for $i = 1, \ldots, m$. It is a fact that whatever physical predictions one makes about a quantum system, they will depend on the $|\psi_i\rangle$ and on the p_i or w_i involved in mixed states or superpositions only through the corresponding matrix ρ. Since the representation of ρ as a mixture of pure states and the representation of a pure state as a superposition of other pure states are highly non-unique, we draw the conclusion that very different ways of preparing a quantum system, which result in the state ρ, cannot be distinguished from one another by any measurement whatsoever on the quantum system. This is a most remarkable feature of quantum mechanics, of absolutely non-classical physical nature.

2.1.2. *The Schrödinger Equation*

Typically the state of a particle undergoes an evolution with time under the influence of an external field. The most basic type of evolution is that of an arbitrary initial state ρ_0 under the influence of a field with Hamiltonian H. This takes the form

$$\rho_t = e^{tH/i\hbar}\rho_0 e^{-tH/i\hbar},$$

where ρ_t denotes the state at time t, $\hbar$ is Planck's constant, and H is a self-adjoint operator on $\mathcal{H}$. If ρ_0 is a pure state then ρ_t is pure for all t and we can choose unit vectors ψ_t such that $\rho_t = |\psi_t\rangle\langle\psi_t|$ and

$$\psi_t = e^{tH/i\hbar}\psi_0\,. \tag{3}$$

Equation (3) is a solution of the celebrated *Schrödinger equation* $i\hbar(\mathrm{d}/\mathrm{d}t)\psi = H\psi$ or equivalently $i\hbar(\mathrm{d}/\mathrm{d}t)\rho = [H, \rho]$.

2.1.3. Separability and Entanglement

When we study several quantum systems (with Hilbert spaces $\mathcal{H}_1, \ldots, \mathcal{H}_m$) interacting together, the natural model for the combined system has as its Hilbert space the tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_m$. Then a state such as $\rho_1 \otimes \cdots \otimes \rho_m$ represents 'particle 1 in state ρ_1 and ... and particle m in state ρ_m'. Suppose the states ρ_i are pure with state-vectors ψ_i. Then the product state we have just defined is also pure with state-vector $\psi_1 \otimes \cdots \otimes \psi_m$. According to the superposition principle, a complex superposition of such state vectors is also a possible state-vector of the interacting systems. Pure states which cannot be written in the product form $\rho_1 \otimes \cdots \otimes \rho_m$ are called *entangled*. The same term is used for mixed states which cannot be written as a mixture of pure product states. A state which is not entangled, is called *separable*. The existence of entangled states is responsible for extraordinary quantum phenomena, which scientists are only just starting to harness (in quantum communication, computation, teleportation, etc.; see Section 8 for an introduction).

An important physical feature of unitary evolution in a tensor product space is that, in general, it does not preserve non-entangledness of states. Suppose that the state $\rho_1 \otimes \rho_2$ evolves according to the Schrödinger operator $U_t = e^{tH/i\hbar}$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$. In general, if H does not have the special form $H_1 \otimes \mathbf{1}_2 + \mathbf{1}_1 \otimes H_2$, the corresponding state at any non-zero time is entangled. The notorious *Schrödinger Cat*, see Section 8.4, is a consequence of this phenomenon of entanglement. For an illustrative discussion of this see, for instance, Isham (1995, Sect. 8.4.2).

2.1.4. Spin-j

So far, our concrete examples have had a two-dimensional Hilbert space. Quantum systems in which the Hilbert space $\mathcal{H}$ is finite-dimensional are sometimes called *spin systems*. A *spin-j system*, where j is a positive half-integer, is one for which the Hilbert space is $\mathbb{C}^{2j+1}$. A physical interpretation of a spin-j system is in terms of a particle having spin angular momentum j.

An important class of spin-j systems can be obtained from pure spin-half systems as follows. Let $|\psi\rangle$ be a state vector representing a spin-half particle in a pure state ρ. Then the quantum system consisting of n independent particles, all prepared in this state, is represented by the state vector $\otimes^n|\psi\rangle$ in $\otimes^n\mathbb{C}^2$. Such state vectors lie in (and span) the subspace

$$\odot^n\mathbb{C}^2 = \mathrm{span}\{\otimes^n|\psi\rangle : |\psi\rangle \in \mathbb{C}^2\}$$

of $\otimes^n\mathbb{C}^2$. The corresponding states have the form $\otimes^n\rho$ and are sometimes known as *(angular momentum) coherent spin-j states*.

Let $\{|\psi_0\rangle, |\psi_1\rangle\}$ be any basis of $\mathbb{C}^2$. Put $j = n/2$ and, for $m = -j, \ldots, j$, define $|m\rangle$ in $\odot^n\mathbb{C}^2$ by

$$|m\rangle = 2^{-j}\Pi_{\odot}\left(\sum_{k=0}^{n}\left(\otimes^k|\psi_0\rangle\right)\otimes\left(\otimes^{n-k}|\psi_1\rangle\right)\right), \tag{4}$$

where $\Pi_{\odot}$ denotes the orthogonal projection from $\otimes^n\mathbb{C}^2$ to $\odot^n\mathbb{C}^2$. The formula

$$\otimes^{2j}\left(\alpha|\psi_0\rangle + \beta|\psi_1\rangle\right) = \sum_{m=-j}^{j}\binom{2j}{m}\alpha^{j+m}\beta^{j-m}|m\rangle \qquad \alpha, \beta \in \mathbb{C}$$

(which can be obtained by binomial expansion) shows that $\{|m\rangle : m = -j, \ldots, j\}$ spans $\odot^n\mathbb{C}^2$. It is easy to check that this is a basis, and so $\odot^n\mathbb{C}^2$ has dimension $2j + 1$.

Example 2 (Coherent Spin-1 states) Take $j = 2$. Then $\{|\psi_0\rangle \otimes |\psi_0\rangle, (|\psi_0\rangle \otimes |\psi_1\rangle + |\psi_1\rangle \otimes |\psi_0\rangle)/\sqrt{2}, |\psi_1\rangle \otimes |\psi_1\rangle\}$ is a basis of $\odot^2\mathbb{C}^2$. Thus $\odot^2\mathbb{C}^2$ can be identified with $\mathbb{C}^3$, whereas $\otimes^2\mathbb{C}^2$ can be identified with $\mathbb{C}^4$. The subspace of $\otimes^2\mathbb{C}^2$ orthogonal to $\odot^2\mathbb{C}^2$ is spanned by $(|\psi_0\rangle \otimes |\psi_1\rangle - |\psi_1\rangle \otimes |\psi_0\rangle)/\sqrt{2}$. The corresponding state is known as the *singlet* or *Bell* state and helps to demonstrate non-classical properties of quantum mechanics; see Section 8.2.

Spin-1 coherent states can be described in matrix terms as follows. If $\rho = \frac{1}{2}(\mathbf{1} + u_x\sigma_x + u_y\sigma_y + u_z\sigma_z)$ is a pure state on $\mathbb{C}^2$ then $u_x^2 + u_y^2 + u_z^2 = 1$ and

$$\rho \otimes \rho = \frac{1}{4}\left\{\mathbf{1} + 2(u_x S_x + u_y S_y + u_z S_z) + u_x^2\sigma_x \odot \sigma_x + u_y^2\sigma_y \odot \sigma_y + u_z^2\sigma_z \odot \sigma_z\right\},$$

where, in terms of the basis $\{|-\frac{1}{2}\rangle, |0\rangle, |\frac{1}{2}\rangle\}$ of $\odot^2\mathbb{C}^2$,

$$S_x = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad S_y = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad S_z = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

and

$$\sigma_x \odot \sigma_x = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \sigma_y \odot \sigma_y = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad \sigma_z \odot \sigma_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

□

2.2. Measurements

Operator-valued probability measures, or OProM's, were introduced in the beginning of the present section. We shall denote by $\mathrm{OProM}(\mathcal{X}, \mathcal{H})$ the set of OProM's on $\mathcal{X}$.

As indicated earlier, a basic kind of operator-valued probability measures consists of those in which the operators $M(A)$ are orthogonal projections. Specifically, a *projector-valued probability measure* (or *PProM*, also called a *simple measurement*) is an operator-valued probability measure M such that

$$M(A) = M(A)^* = M(A)^2 \qquad A \in \mathcal{A}.$$

We shall denote by $\mathrm{PProM}(\mathcal{X}, \mathcal{H})$ the set of PProM's on $\mathcal{X}$. As we noted, when the outcome space $\mathcal{X}$ is $\mathbb{R}$, the PProM's stand in one-to-one correspondence with the self-adjoint operators on $\mathcal{H}$, which in this context are also called observables. If one measures the observable X on a quantum system in state ρ, it turns out that the expected value of the outcome is given by the *trace rule*

$$\mathrm{E}(\mathrm{meas}(X;\rho)) = \sum_x x \,\mathrm{tr}\{\rho\Pi_{[x]}\} = \mathrm{tr}\{\rho \sum_x x\Pi_{[x]}\} = \mathrm{tr}\{\rho X\}. \tag{5}$$

Example 3 (Spin-half, cont.) For any unit vector ψ of $\mathbb{C}^2$, the observable $2|\psi\rangle\langle\psi| - \mathbf{1} = |\psi\rangle\langle\psi| - |\psi^\perp\rangle\langle\psi^\perp|$ defines a PProM. It has eigenvalues 1 and -1 and one-dimensional eigenspaces spanned by ψ and $\psi^\perp$. This operator measures the spin of the particle in the direction (on the Poincaré sphere) defined by ψ. We mentioned two of such measurements in Section 2.1.1 on mixing and superposition. □

Example 4 (Spin-half, cont.) In particular, with $\mathcal{X} = \{-1, 1\}$, the specification

$$\begin{aligned} M(\{+1\}) &= \tfrac{1}{2}(\mathbf{1} + \sigma_x) \\ M(\{-1\}) &= \tfrac{1}{2}(\mathbf{1} - \sigma_x) \end{aligned}$$

defines an element of $\mathrm{PProM}(\mathcal{X}, \mathbb{C}^2)$. It corresponds to the observable σ_x: spin in the x-direction. □

We next discuss the notion of *quantum randomisation* whereby adding an auxiliary quantum system to a system under study gives one further possibilities for probing the system of interest. This also connects to the important notion of *realisation*: representing generalised measurements by simple measurements on a quantum randomised extension.

Suppose given a Hilbert space $\mathcal{H}$, and a pair $(\mathcal{K}, \rho_a)$, where $\mathcal{K}$ is a Hilbert space and ρ_a is a state on $\mathcal{K}$. Any measurement $\widetilde{M}$ in $\mathrm{OProM}(\mathcal{X}, \mathcal{H} \otimes \mathcal{K})$ induces a measurement M in $\mathrm{OProM}(\mathcal{X}, \mathcal{H})$ which is determined by

$$\mathrm{tr}\{\rho M(A)\} = \mathrm{tr}\left\{(\rho \otimes \rho_a)\widetilde{M}(A)\right\} \quad \rho \in \mathcal{S}(\mathcal{H}), \quad A \in \mathcal{A}. \tag{6}$$

The pair $(\mathcal{K}, \rho_a)$ is called an *ancilla*. The following theorem (Holevo's extension of Naimark's Theorem, see Appendix A.2) states that any measurement M in $\mathrm{OProM}(\mathcal{X}, \mathcal{H})$ is of the form (6) for some ancilla $(\mathcal{K}, \rho_a)$ and some simple measurement $\widetilde{M}$ in $\mathrm{PProM}(\mathcal{X}, \mathcal{H} \otimes \mathcal{K})$. The triple $(\mathcal{K}, \rho_a, \widetilde{M})$ is called a *realisation* of M (the words *extension* or *dilation* are also used sometimes). Adding an ancilla before taking a simple measurement could be thought of as *quantum randomisation*.

Theorem 1 (Holevo 1982) *For every M in* $\mathrm{OProM}(\mathcal{X}, \mathcal{H})$*, there is an ancilla $(\mathcal{K}, \rho_a)$ and an element $\widetilde{M}$ of* $\mathrm{PProM}(\mathcal{X}, \mathcal{H} \otimes \mathcal{K})$ *which form a realisation of M.*

We use the term quantum randomisation, because of its analogy with the mathematical representation of randomisation in classical statistics, whereby one replaces the original probability space with a product space, one of whose components is the original space of interest, while the other corresponds to an independent random experiment with probabilities under the control of the experimenter. Just as randomisation in classical statistics is sometimes needed to solve optimisation problems of statistical decision theory, quantum randomisation sometimes allows for strictly better solutions than can be obtained without it.

Here is a simple spin-half example of an OProM which cannot be represented without *quantum* randomisation.

Example 5 (The triad) The triad, or Mercedes-Benz logo, has an outcome space consisting of just three outcomes: let us call them 1, 2 and 3. Let $\vec{v}_i$, $i = 1, 2, 3$, denote three unit vectors in the same plane through the origin in R^3, at angles of 120° to one another. Then the matrices $M(\{i\}) = \frac{1}{3}(\mathbf{1} + \vec{v}_i \cdot \vec{\sigma})$ define an OProM on the sample space $\{1, 2, 3\}$. It turns up as the optimal solution to the decision problem: suppose a spin-half system is generated in one of the three states $\rho_i = \frac{1}{2}(\mathbf{1} - \vec{v}_i \cdot \vec{\sigma})$, $i = 1, 2, 3$, with equal probabilities. What decision rule gives the maximum probability of guessing the actual state correctly? There is no way to equal the success probability of this method, if one uses only simple measurements, even allowing for (classically) randomised procedures. □

Finally, we introduce some further terminology concerning measurements. Given an OProM M and a measurable function T from its outcome space $\mathcal{X}$ to another space $\mathcal{Y}$, one can define a new measurement $M' = M \circ T^{-1}$ with outcome space $\mathcal{Y}$. It corresponds to restricting attention to the function T of the outcome of the first measurement M. We call it a *coarsening* of the original measurement, and conversely we say that M is a *refinement* of M'.

A measurement M is called *dominated* by a (real, sigma-finite) measure ν on the outcome space, if there exists a non-negative self-adjoint matrix-valued function $m(x)$, called the density of M, such that $M(B) = \int_B m(x)\nu(\mathrm{d}x)$ for all B. In the finite-dimensional case every measurement is dominated: take $\nu(B) = \mathrm{trace}(M(B))$.

To exemplify these notions, suppose for some dominated measurement M one can write $m(x) = m_1(x) + m_2(x)$ for two non-negative self-adjoint matrix-valued functions m_1 and m_2. Then one can define a refinement M' of M as the measurement on the outcome space $\mathcal{X}' = \mathcal{X} \times \{1, 2\}$ with density $m_i(x)$, $(x, i) \in \mathcal{X}'$, with respect to the product of ν with counting measure.

We described earlier how one can form product spaces from separate quantum systems, leading to notions of product states, separable states, and entangled states. Given an OProM M on one component of a product space, one can naturally talk about 'the same measurement' on the product system. It has components $M(B) \otimes \mathbf{1}$. Given measurements M and M' defined on the two components of a product system, one can define in a natural way the measurement 'apply simultaneously M and M' to each component': its outcome space is the product of the two outcome spaces, and it is defined using obvious notation by $M \otimes M'(B \times B') = M(B) \otimes M'(B')$.

A measurement M on a product space is called *separable* if it has a density m such that each $m(x)$ can be written as a positive linear combination of tensor products of non-negative components. It can then be thought of as a coarsening of a measurement with density m' such that each $m'(y)$ is a product of non-negative components.

2.3. Instruments

When a physical measurement is made on a quantum system, the system usually changes state in some stochastic manner. Thus a complete description of the measurement specifies not just the probability distribution of the outcome x but also the new state of the system when the outcome is x. We shall refer to the states of the system before and after measurement as the *prior state* and the *posterior state*, and use the notation $\mathcal{N}$ to denote a particular mapping from prior states to probability distributions over outcomes, with a particular posterior state associated with each outcome and given prior state. Such mappings are called *instruments* (Davies and Lewis 1970; Davies 1976). Because of the basic rules of quantum mechanics, an instrument cannot be completely arbitrary but must satisfy certain constraints. We shall describe these constraints after we have introduced some further notation.

The word 'instrument' is not very illuminating. The concept which we are trying to catch here is that of any interaction between a quantum system and the real world. The interaction will change the state of the quantum system, and cause changes in the real world. One can think of these changes as being information recorded in classical physical systems. Data stored on a CD-ROM or printed on paper is just one kind of classical physical information. A measurement, in the sense of a deliberately carried out experiment, is just one kind of interaction. The data which are available to an experimenter, after a measurement has been done, form only part of the totality of changes which have happened in the real world. So one can distinguish between what is somehow imprinted in the real world as a result of the interaction which takes place when the instrument is applied to the quantum system, and a coarsened or reduced version of this information, which is the outcome of the measurement as it is available to the experimenter. What is relevant for the experimenter is the final state of the quantum system, conditioned on the data which he has available. This is typically different from the final state of the quantum system, conditioned on the final state of the real world.

In the following, the outcome of the instrument will refer to the data available to the experimenter, and the posterior state means the final state (possibly mixed) of the quantum system given this information only.

Consider an instrument $\mathcal{N}$ with outcomes x in the measurable space $(\mathcal{X}, \mathcal{A})$. Let $\pi(\mathrm{d}x; \rho, \mathcal{N})$ denote the probability distribution of the outcome of the measurement, and let $\sigma(x; \rho, \mathcal{N})$ denote the posterior state when the prior state is ρ and the outcome of the measurement is x. Now let Y denote some observable on the quantum system and let $A \in \mathcal{A}$ denote a measurable set of outcomes. Suppose one 'measures the instrument' on the state ρ, registers whether or not the outcome is in A, and subsequently measures the observable Y. Then the expected value of the indicator of the event 'outcome is in A' times the outcome of measuring the observable Y is the number $\int_A \pi(\mathrm{d}x; \rho, \mathcal{N})\mathrm{tr}\{(\sigma(x; \rho, \mathcal{N})Y\}$, by using the trace rule (5). Now it turns out that this number, seen as a function of prior state ρ, measurable subset of outcomes A, and observable Y, determines $\mathcal{N}$ completely. By the interpretation of mixed states as probability mixtures, it follows that the expression is linear in ρ and therefore can be rewritten as $\mathrm{tr}\{\rho\mathcal{N}(A)[Y]\}$ where $\mathcal{N}(A)[Y]$, for each event A in the outcome space and each observable Y, is a uniquely defined (possibly unbounded) self-adjoint operator on $\mathcal{H}$. This linearity constraint restricts considerably the class of all possible (π, σ). One can show that $\mathcal{N}(A)[Y]$ must be countably additive in the argument A, linear and positive in Y (positive in the sense of mapping nonnegative operators to nonnegative operators), and normalised in the sense that $\mathcal{N}(\mathcal{X})[\mathbf{1}] = \mathbf{1}$.

Thus, mathematically, an instrument $\mathcal{N}$ can be specified equally well by giving the probability distribution of the outcome of the measurement $\pi(\mathrm{d}x; \rho, \mathcal{N})$, together with the posterior state $\sigma(x; \rho, \mathcal{N})$, as by giving an operator $\mathcal{N}(A)[Y]$ for each A and Y. The physical constraints imposed by quantum theory restrict the possible (π, σ), and equivalently restrict the possible $\mathcal{N}(A)[Y]$. The second specification is less direct but more convenient from a theoretical point of view, since the physical constraints (additivity, linearity, positivity, normalization) are much more simple to express in those terms. In a moment we indicate that, on further physical considerations, the positivity condition should be strengthened to a condition called *complete positivity*.

Following Ozawa (1985), we show how to recover (π, σ) from $\mathcal{N}$. The first step is to read off the measurement or OProM M which is determined by the instrument $\mathcal{N}$, when we ignore the posterior state. This is given by the prescription

$$M(A) = \mathcal{N}(A)[\mathbf{1}]\,. \tag{7}$$

The probability that the measurement of the state ρ results in an outcome in A is given by

$$\pi(A; \rho, \mathcal{N}) = \mathrm{tr}\{\rho\mathcal{N}(A)[\mathbf{1}]\}.$$

If the system was in state ρ just before the measurement then the state of the system after the measurement, given that the measurement was observed to result in an outcome belonging to A, is determined as the solution $\sigma(A; \rho, \mathcal{N})$ of the equation

$$\mathrm{tr}\{\sigma(A; \rho, \mathcal{N})Y\} = \frac{\mathrm{tr}\{\rho\mathcal{N}(A)[Y]\}}{\mathrm{tr}\{\rho\mathcal{N}(A)[\mathbf{1}]\}} \qquad Y \in \mathbb{B}(\mathcal{H})$$

(provided that $\mathrm{tr}\{\rho\mathcal{N}(A)[\mathbf{1}]\} > 0$). Finally, the family $\sigma(x; \rho, \mathcal{N})$ of posterior states is characterised (almost everywhere, with respect to π) by

$$\mathrm{tr}\{\rho\mathcal{N}(A)[Y]\} = \int_A \mathrm{tr}\{\sigma(x; \rho, \mathcal{N})Y\}\pi(\mathrm{d}x; \rho, \mathcal{N}) \qquad Y \in \mathbb{B}(\mathcal{H}) \qquad A \in \mathcal{A}\,.$$

An extremely important class of quantum instruments consists of those of the form

$$\mathcal{N}(\mathrm{d}x)[Y] = \sum_i W_i(x)^* Y W_i(x)\nu(\mathrm{d}x)\,, \tag{8}$$

where ν is a σ-finite measure on $\mathcal{X}$ (and, without loss of generality, can be taken to be a probability measure), the index i runs over some finite or countable set, and W_i is a measurable function from $\mathcal{X}$ to $\mathbb{B}(\mathcal{H})$ such that

$$\sum_i \int_{\mathcal{X}} W_i(x) W_i(x)^* \nu(\mathrm{d}x) = \mathbf{1}\,.$$

For such quantum instruments, the posterior states are

$$\sigma(x;\rho,\mathcal{N}) = \frac{\sum_i W_i(x)^* \rho W_i(x)}{\sum_i \mathrm{tr}\{\rho W_i(x) W_i(x)^*\}}$$

and the distribution of the outcome is

$$\pi(\mathrm{d}x;\rho,\mathcal{N}) = \sum_i \mathrm{tr}\{\rho W_i(x) W_i(x)^*\}\nu(\mathrm{d}x).$$

Such quantum instruments are almost generic, in the sense that an instrument which satisfies the further physically motivated condition of *complete positivity* can be represented as in (8), except that the operators $W_i(x)$ need not be bounded (in which case the formulae we have given need to be interpreted with some care).

The mathematical definition of complete positivity is given in Appendix A.1. Its intuitive meaning is as follows. We can consider the instrument as acting not just on the system of interest $\mathcal{H}$ but also on a completely arbitrary system $\mathcal{K}$ somewhere else in the universe. If the systems are independent, we can express the joint state as a tensor product, and the instrument acts on it by transforming the system of interest as we have already specified, while leaving the auxiliary system unchanged; the posterior joint state remains a product state. Now once we have specified how the extended instrument acts on product states, one can calculate how it acts on *any* joint state, including entangled states, by using the linearity which is a basic feature of quantum physics. To be physically meaningful, this extended instrument has to be positive, in the sense of mapping states (nonnegative matrices) to states (after all, the system we are studying may actually be in an entangled state with a system elsewhere). The mathematical statement of this physical property is called *complete positivity*.

Formulae like (8) are known in the physics literature as Kraus representations. If we allow *unbounded instruments* for which the self-adjoint operator $\mathcal{N}(A)[Y]$ is not necessarily bounded for all A and Y, then the $W_i(x)$ need not be bounded either. In this case posterior states may not be defined for each outcome of the measurement, but only for each measurable collection of outcomes of positive probability. Allowing unbounded operators as well as bounded makes a difference only in infinite dimensional spaces, see Example 15 in Appendix A.1 Key references on instruments and complete positivity are Stinespring (1955), Davies and Lewis (1970), Davies (1976), Kraus (1983), Ozawa (1985), Loubenets (1999, 2000), and Holevo (2001c).

Example 6 (Simple Instruments) Let $\{\Pi_{[x]} : x \in \mathcal{X}\}$ define a PProM on a finite-dimensional quantum system, corresponding to the simple measurement of the observable $Q = \sum x \Pi_{[x]}$. One can embed this measurement in many different instruments, i.e., the state could be transformed by the measurement in many different ways. However the most simple description possible is obtained when one takes, in (8), ν to be counting measure on the finite set $\mathcal{X}$, the set of indices i to contain a single element, and $W_i(x) = W(x) = \Pi_{[x]}$. We call this particular instrument the corresponding *simple* instrument. If one applies it to a system in the pure state with state-vector ψ, and observes the outcome x, then the state of the system remains pure but now has state vector $\Pi_{[x]}\psi/\|\Pi_{[x]}\psi\|$. The probability of this event is precisely $\|\Pi_{[x]}\psi\|^2$. When the state transforms in this way, one says that von Neumann's or Lüders' *projection postulate* holds for the measurement of the observable Q.

Two observables Q, P are called *compatible* if as operators they commute. For a Borel measurable function $f : \mathbb{R} \to \mathbb{R}$ and an observable R with eigenvalues r and eigenspaces the ranges of the projectors $\Pi_{[R=r]}$, the observable $f(R)$ is the operator $\sum f(r)\Pi_{[R=r]}$. A celebrated result of von Neumann is that observables Q and P are compatible if and only if they are both functions $f(R)$, $g(R)$ of a third observable R. Taking R to have as coarse a collection of eigenspaces as possible, one can show that the results of the following three instruments are identical: the simple instrument for Q followed by the simple instrument for P, recording the values q of Q and p of P; the simple instrument for P followed by the simple instrument for Q, recording the values q of Q and p of P; and the simple instrument for R, recording the values $q = f(r)$ and $p = g(r)$ where r is the observed value of R.

It follows that the probability distribution of the outcome of measurement of an observable P is not altered when it is measured (simply, jointly) together with any other compatible observables. Note that the expected value of the outcome of a measurement of the observable Q on a quantum system in state ρ is $\mathrm{tr}\{\rho Q\}$, and the expected value of the real function f of this outcome is $\mathrm{tr}\{\rho f(Q)\}$, identical to the expectation of the outcome of a measurement of the observable $f(Q)$. We call this rule *the law of the unconscious quantum physicist* since it is analogous to the law of the unconscious statistician, according to which the expectation of a function $Y = f(X)$ of a random variable X may be calculated by an integration (i) over the underlying probability space, (ii) over the outcome space of X, (iii) over the outcome space of Y.

A useful consequence of this calculus of functions of observables is that the characteristic function of the distribution of a measurement of an observable Q is equal to $\mathrm{tr}\{\rho e^{itQ}\}$. Since Q is self-adjoint, e^{itQ} is unitary and the trace may have a physical interpretation which aids its calculation. □

Further results of Ozawa (1985) generalise the realisability of measurements (Naimark, Holevo theorems) to the realisability of an arbitrary completely positive instrument. Namely, after forming a compound system by taking the tensor product with some ancilla, the instrument can be realised as a unitary (Schrödinger) evolution for some length of time, followed by the action of a simple instrument (measurement of an observable, with state transition according to von Neumann's projection postulate). Therefore to say that the most general operation on a quantum system is a completely positive instrument comes down to saying: the only mechanisms known in quantum mechanics are Schrödinger evolution, von Neumann measurement, and forming compound systems (entanglement). Combining these ingredients in arbitrary ways, one remains within the class of completely positive instruments; moreover, anything in that class can be realised in this way.

Just as we introduced notions of coarsening and refinement for OProM's, and discussed OProM's on product systems, one can do the same (and more) for instruments. The extra ingredient is *composition*. Since the description of an instrument includes the state of the system after the measurement by the instrument, we are able to define mathematically the composition of two instuments, corresponding to the notion of applying first one instrument, and then the second, while registering the outcomes (data) produced at each of the two stages. The outcome space of the composition of two instruments is the product of the two respective outcome spaces. A more complicated form of composition is possible, in which the second instrument is replaced by a family of instruments, indexed by possible outcomes of the first instrument. Informally: apply the first instrument, then choose a second instrument depending on the outcome of the first; keep the outcomes of both. We do not write out the mathematical formalism for describing these rather natural concepts.

For coarsening, we do write out some formal details, since we need later to refer to a specific result. Let $\mathcal{N}$ denote an instrument on a Hilbert space $\mathcal{H}$ and with outcome space $(\mathcal{X}, \mathcal{A})$ and let $\mathcal{N}'$ be a *coarsening* of $\mathcal{N}$, i.e. $\mathcal{N}'$ is an instrument on the same Hilbert space $\mathcal{H}$, with outcome space $(\mathcal{Y}, \mathcal{B})$, and there is a mapping T from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ such that

$$\mathcal{N}'(B)[\cdot] = \mathcal{N}(T^{-1}(B))[\cdot]$$

for all $B \in \mathcal{B}$. This mathematical formalism defines the instrument corresponding to applying the instrument $\mathcal{N}$, registering the result of applying the function T to the outcome x, and discarding x. Because of this interpretation, one has the following relation between the posterior states $\sigma(x; \rho, \mathcal{N})$ and $\sigma(t; \rho, \mathcal{N}')$:

$$\sigma(t; \rho, \mathcal{N}') = \int_{T^{-1}(t)} \sigma(x; \rho, \mathcal{N})\pi(\mathrm{d}x|t; \rho, \mathcal{N}), \tag{9}$$

where $\pi(\mathrm{d}x|t; \rho, \mathcal{N})$ is the conditional distribution of x given $T(x) = t$ computed from $\pi(\mathrm{d}x; \rho, \mathcal{N})$.

An instrument defined on one component of a product system can be extended in a natural way (similar to that described in Section 2.2 for measurements) to an instrument on the product system. Conversely, it is of great interest whether instruments on a product system can in some way be reduced to 'separate instruments on the separate sub-systems'. There are two important notions in this context. The first (similar to the concept of separability of measurements) is the *mathematical* concept of separability of an instrument defined on a product system: this is that each $W_i(x)$ in some representation (8) is a tensor product of separate matrices for each component. The second is the *physical* property which we shall call *multilocality*: an instrument is called multilocal, if it can be represented as a coarsening of a composition of separate instruments applied sequentially to separate components of the product system, where the choice of each instrument at each stage may depend on the outcomes of the instruments applied previously. Moreover, each component of the system may be measured several times (i.e., at different stages), and the choice of component measured at the nth stage may depend on the outcomes at previous stages. One should think of the different components of the quantum system as being localised at different locations in space. At each location separately, anything quantum is allowed, but all communication between locations is classical. It is a theorem of Bennett et al. (1999a) that every multilocal instrument is separable, but that (surprisingly) not all separable instruments are multilocal. It is an open problem to find a physically meaningful characterisation of separability, and conversely to find a mathematically convenient characterisation of multilocality. (Note, our terminology is not standard: the word 'unentangled' is used by some authors instead of separable, and 'separable' instead of multilocal).

Not all joint measurements (by which we just mean instruments on product systems), are separable, let alone multilocal. Just as quantum randomised measurements can give strictly more powerful ways to probe the state of a quantum system than (combinations of) simple measurements and classical randomisation, so non-separable measurements can do strictly better than separable measurements at extracting information from product systems, even if a priori there is no interaction of any kind between the subsystems; this is a main conclusion of Section 6.3.

3. Parametric Quantum Models and Likelihood

A measurement from a parametric quantum model $(\boldsymbol{\rho}, m)$ results in an observation x with density

$$p(x;\theta) = \mathrm{tr}\{\rho(\theta)m(x)\}$$

and log likelihood

$$l(\theta) = \log \mathrm{tr}\{\rho(\theta)m(x)\}\,.$$

For simplicity, let us suppose θ is one-dimensional. For the calculation of log likelihood derivatives in the present setting it is convenient to work with the *symmetric logarithmic derivative* or *quantum score* of $\boldsymbol{\rho}$, denoted by $\rho_{/\!\!/\theta}$. This is defined implicitly as the self-adjoint solution of the equation

$$\rho_{/\theta} = \rho \circ \rho_{/\!\!/\theta}\,, \tag{10}$$

where $\circ$ denotes the Jordan product, i.e.

$$\rho \circ \rho_{/\!\!/\theta} = \tfrac{1}{2}(\rho\rho_{/\!\!/\theta} + \rho_{/\!\!/\theta}\rho)\,,$$

$\rho_{/\theta}$ denoting the ordinary derivative of ρ with respect to θ (term by term differentiation in matrix representations of ρ). (We shall often suppress the argument θ in quantities like ρ, $\rho_{/\theta}$, $\rho_{/\!\!/\theta}$, etc.) The quantum score exists and is essentially unique subject only to mild conditions (for a discussion of this see, for example, Holevo 1982).

The likelihood score $l_{/\theta}(\theta) = (\mathrm{d}/\mathrm{d}\theta)l(\theta)$ may be expressed in terms of the quantum score $\rho_{/\!\!/\theta}(\theta)$ of $\rho(\theta)$ as

$$\begin{aligned} l_{/\theta}(\theta) &= p(x;\theta)^{-1}\mathrm{tr}\{\rho_{/\theta}(\theta)m(x)\} \\ &= p(x;\theta)^{-1}\tfrac{1}{2}\mathrm{tr}\,\{(\rho(\theta)\rho_{/\!\!/\theta}(\theta) + \rho_{/\!\!/\theta}(\theta)\rho(\theta))m(x)\} \\ &= p(x;\theta)^{-1}\Re\,\mathrm{tr}\,\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)m(x)\}\,, \end{aligned}$$

where we have used the fact that for any self-adjoint operators P, Q, R on $\mathscr{H}$ the trace operation satisfies $\mathrm{tr}\{PQR\} = \overline{\mathrm{tr}\{RQP\}}$ and $\Re\,\mathrm{tr}\{Q\} = \frac{1}{2}\mathrm{tr}\{Q + Q^*\}$. It follows that

$$\mathrm{E}_\theta[l_{/\theta}(\theta)] = \mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)\}\,.$$

Thus, since the mean value of $l_{/\theta}$ is 0, we find that

$$\mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)\} = 0\,. \tag{11}$$

The expected (Fisher) information $i(\theta) = i(\theta; M) = \mathrm{E}_\theta[l_{/\theta}(\theta)^2]$ may be written as

$$i(\theta; M) = \int p(x;\theta)^{-1}\left\{\Re\,\mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)m(x)\}\right\}^2 \nu(\mathrm{d}x)\,. \tag{12}$$

It plays a key role in the quantum context, just as in classical statistics, and is discussed in Section 6. In particular, we will there discuss its relation with the expected or Fisher *quantum information*

$$I(\theta) = \mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)^2\}. \tag{13}$$

The quantum score is a self-adjoint operator, and therefore may be interpreted as an observable which one might measure on the quantum system. What we have just seen is that the outcome of a simple measurement of the quantum score has mean zero, and variance equal to the quantum Fisher information.

4. Quantum Exponential and Quantum Transformation Models

In traditional statistics, the two major classes of parametric models are the exponential models (in which the log densities are affine functions of appropriate parameters) and the transformation (or group) models (in which a group acts in a consistent fashion on both the sample space and the parameter space); see Barndorff-Nielsen and Cox (1994). The intersection of these classes is the class of exponential transformation models, and its members have a particularly nice structure. There are quantum analogues of these classes, and they have useful properties.

4.1. Quantum Exponential Models

A *quantum exponential model* is a quantum statistical model for which the states $\rho(\theta)$ can be represented in the form

$$\rho(\theta) = e^{-\kappa(\theta)} e^{\frac{1}{2}\overline{\gamma}^r(\theta)T_r^*} \rho_0 e^{\frac{1}{2}\gamma^r(\theta)T_r} \qquad \theta \in \Theta\,,$$

where $\gamma = (\gamma^1, \ldots, \gamma^k) : \Theta \to \mathbb{C}^k$, $T_1, \ldots, T_k$ are operators on $\mathcal{H}$, ρ_0 is self-adjoint and non-negative (but not necessarily a density matrix), the Einstein summation convention (of summing over any index which appears as both a subscript and a superscript) has been used, and $\kappa(\theta)$ is a log norming constant, given by

$$\kappa(\theta) = \log \operatorname{tr}\{e^{\frac{1}{2}\overline{\gamma}^r(\theta)T_r^*} \rho_0 e^{\frac{1}{2}\gamma^r(\theta)T_r}\}\,.$$

Three important special types of quantum exponential model are those in which $T_1, \ldots, T_k$ are bounded and self-adjoint, (and for the first type, T_0, $T_1, \ldots, T_k$ all commute) and the quantum states have the forms

$$\rho(\theta) = e^{-\kappa(\theta)} \exp\left\{T_0 + \theta^r T_r\right\} \tag{14}$$

$$\rho(\theta) = e^{-\kappa(\theta)} \exp\left\{\tfrac{1}{2}\theta^r T_r\right\} \rho_0 \exp\left\{\tfrac{1}{2}\theta^r T_r\right\} \tag{15}$$

$$\rho(\theta) = \exp\left\{-i\tfrac{1}{2}\theta^r T_r\right\} \rho_0 \exp\left\{i\tfrac{1}{2}\theta^r T_r\right\}\,, \tag{16}$$

respectively, where $\theta = (\theta^1, \ldots, \theta^k) \in \mathbb{R}^k$ and $\rho_0 \in \mathbb{SA}_+(\mathcal{H})$, and the summation convention is in force.

We call these three types, the quantum exponential models of *mechanical* type, *symmetric* type, and *unitary* type respectively. The mechanical type arises (at least, with $k = 1$) in quantum statistical mechanics as a state of statistical equilibrium, see Gardiner and Zoller (2000, Sect. 2.4.2). The symmetric type has theoretical statistical significance, as we shall see, connected among other things to the fact that the quantum score for this model is easy to compute explicitly. The unitary type has physical significance connected to the fact that it is also a transformation model (quantum transformation models are defined in the next subsection). The mechanical type is a special case of the symmetric type when T_0, $T_1, \ldots, T_k$ all commute.

In general, the statistical model obtained by applying a measurement to a quantum exponential model is not an exponential model (in the classical sense). However, for a quantum exponential model of the form (15) in which

$$T_j = t_j(X) \quad j = 1, \ldots, k \qquad \text{for some } X \text{ in } \mathbb{SA}(\mathcal{H})\,, \tag{17}$$

i.e., the T_j commute, the statistical model obtained by applying the measurement X is a full exponential model. Various pleasant properties of such quantum exponential models then follow from standard properties of the full exponential models.

The classical Cramér–Rao bound for the variance of an unbiased estimator t of θ is

$$\mathrm{Var}(t) \geq i(\theta; M)^{-1} . \tag{18}$$

Combining (18) with Braunstein and Caves' (1994) quantum information bound $i(\theta; M) \leq I(\theta)$, which we derive as (31) in Section 6.2, yields Helstrom's (1976) quantum Cramér–Rao bound

$$\mathrm{Var}(t) \geq I(\theta)^{-1} , \tag{19}$$

whenever t is an unbiased estimator based on a quantum measurement. It is a classical result that, under certain regularity conditions, the following are equivalent: (i) equality holds in (18), (ii) the score is an affine function of t, (iii) the model is exponential with t as canonical statistic (cf. pp. 254–255 of Cox and Hinkley 1974). This result has a quantum analogue, see Theorems 3 and 4 and Corollary 1 below, which states that under certain regularity conditions, there is equivalence between (i) equality holds in (19) for some unbiased estimator t based on some measurement M, (ii) the symmetric quantum score is an affine function of commuting $T_1, \ldots, T_k$, and (iii) the quantum model is a quantum exponential model of type (15) where $T_1, \ldots, T_k$ satisfy (17). The regularity conditions which we assume below are indubitably too strong: the result should be true under minimal smoothness assumptions.

4.2. Quantum Transformation Models

Consider a parametric quantum model $(\boldsymbol{\rho}, M)$ consisting of a family $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ of states and a measurement M with outcome space $(\mathcal{X}, \mathcal{A})$. Suppose there exists a group, G, with elements g, acting both on $\mathcal{X}$ and on Θ in such a way that the following consistency condition holds

$$\mathrm{tr}\{\rho(\theta)M(A)\} = \mathrm{tr}\{\rho(g\theta)M(g^{-1}A)\} \tag{20}$$

for all θ, A and g. If, moreover, G acts transitively on Θ we say that $(\boldsymbol{\rho}, M)$ is a quantum transformation model. In this case, the resulting statistical model for the outcome of a measurement of M, i.e. $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, where $\mathcal{P} = \mathrm{tr}\{\rho(\theta)M\} : \theta \in \Theta\}$, is a classical transformation model. Consequently, the Main Theorem for transformation models, see Barndorff-Nielsen and Cox (1994, pp. 56–57) and references given there, applies to $(\mathcal{X}, \mathcal{A}, \mathcal{P})$.

Of particular physical interest are situations where the actions of G are such that

$$M(g^{-1}A) = U_g^* M(A) U_g \qquad A \in \mathcal{A}, \tag{21}$$

$$\rho(g\theta) = U_g^* \rho(\theta) U_g , \tag{22}$$

where the U_g are unitary matrices satisfying

$$U_{gh} = w(g, h) U_g U_h \qquad g, h \in G , \tag{23}$$

for some complex valued function w with $|w(g, h)| = 1$ for all g and h. A mapping $g \mapsto U_g$ with the property (23) is said to constitute a *projective unitary representation* of G and a measurement M satisfying (21) is termed *covariant* in the physical literature; *equivariant* would be a more correct terminology. Under certain conditions, equivariant measurements are representable in the form

$$M(A) = \int_{\{g: g^{-1}x_0 \in A\}} U_g^* R_0 U_g \mu(\mathrm{d}g)$$

for an invariant measure μ on G, a fixed non-negative self-adjoint operator R_0 on $\mathcal{H}$ and some fixed point $x_0 \in \mathcal{X}$.

Example 7 (Equivariant measurements for spin-half) Suppose both outcome space $\mathcal{X}$ and group G are the unit circle S^1. Let the Hilbert space $\mathcal{H}$ be $\mathbb{C}^2$ and let S^1 act on $\mathcal{H}$ via the projective representation

$$\phi \mapsto U_\phi = \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix} \qquad \phi \in S^1 .$$

Then by Holevo (1982, p. 175 with $j = \frac{1}{2}$) any equivariant M has

$$m(\phi) = \begin{pmatrix} 1 & a e^{i\phi} \\ \overline{a} e^{-i\phi} & 1 \end{pmatrix}$$

with respect to the uniform distribution on S^1, for some a with $|a| \leq 1$. □

Example 8 (Equivariant measurements for spin-j) The preceding example generalises to spin-j coherent states. Again, both the outcome space $\mathcal{X}$ and the group G are the unit circle S^1. Now let the Hilbert space $\mathcal{H}$ be $\odot^n \mathbb{C}^2$. Define the operator J on $\mathcal{H}$ by

$$J = \sum_{m=-j}^{j} m |m\rangle\langle m| ,$$

where $j = n/2$ and $|m\rangle$ is defined in (4). Then putting

$$U_\phi = e^{i\phi J} \qquad \phi \in S^1$$

gives a projective representation of S^1 on $\mathcal{H}$. By Holevo (1982, p. 175) any equivariant measurement has density

$$m(\phi) = e^{-i\phi J} R_0 e^{i\phi J}$$

with respect to the uniform distribution on S^1, for some positive operator R_0 satisfying

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-i\phi J} R_0 e^{i\phi J} \mathrm{d}\phi = \mathbf{1} .$$

□

4.3. Quantum Exponential Transformation Models

A *quantum exponential transformation model* is a quantum exponential model which is also a quantum transformation model. The pleasant properties of classical exponential transformation models (Barndorff-Nielsen et al., 1982) are shared by a large class of quantum exponential transformation models of the form (15) which satisfy (17). In particular, if $\mathcal{H}$ is finite-dimensional and the group acts transitively then there is a unique affine action of the group on $\mathbb{R}^k$ such that $(t_1, \ldots, t_k) : \mathcal{X} \to \mathbb{R}^k$ is equivariant.

Example 9 (Spin-half: great circle model) Consider the spin-half model $\rho(\theta) = U \frac{1}{2}(\mathbf{1} + \cos\theta\sigma_x + \sin\theta\sigma_y) U^*$ where U is a fixed 2×2 unitary matrix, and σ_x and σ_y are two of the Pauli spin matrices, while the parameter θ varies through $[0, 2\pi)$; see Example 1. The matrix U can always be written as $\exp(-i\phi\vec{u} \cdot \vec{\sigma})$ for some real three-dimensional unit vector $\vec{u}$ and angle ϕ. Considered as a curve on the Poincaré sphere, the model forms a great circle. If U is the identity (or, equivalently, $\phi = 0$) the curve just follows the line of the equator; the presence of U corresponds to rotating the sphere carrying this curve about the direction $\vec{u}$ through an angle ϕ. Thus our model describes an arbitrary great circle on the Poincaré sphere, parametrised in a natural way. Since we can write $\rho(\theta) = UV_\theta U^* \rho(0) UV_\theta^* U^*$, where the unitary matrix V_θ corresponds to rotation of the Poincaré sphere by an angle θ about the z-axis, we can write this model as a unitary transformation model of the form (22). Together with any equivariant measurement, this model forms a quantum transformation model. The model is clearly also an exponential model of unitary type. Perhaps surprisingly, it can be reparameterised so as also to be an exponential model of symmetric type. We leave the details (which depend on the algebraic properties of the Paul spin matrices) to the reader, but just point out that a one-parameter pure-state exponential model of symmetric type has to be of the form $\rho(\theta) = \exp(-\kappa(\theta)) \exp(\frac{1}{2}\theta\vec{u}\cdot\vec{\sigma})\frac{1}{2}(\mathbf{1}+\vec{v}\cdot\vec{\sigma})\exp(\frac{1}{2}\theta\vec{u}\cdot\vec{\sigma})$ for some real unit vectors $\vec{u}$ and $\vec{v}$, since every self-adjoint 2×2 matrix is an affine function of a spin matrix $\vec{u}\cdot\vec{\sigma}$. Now write out the exponential of a matrix as its power series, and use the fact that the square of any spin matrix is the identity.

This example is due to Fujiwara and Nagaoka (1995). □

5. Quantum Exhaustivity and Sufficiency

This section introduces and relates some concepts connected to the classical notion of sufficiency.

5.1. Quantum Exhaustivity

An important role is played by quantum instruments for which no information on the unknown parameter of a quantum parametric model of states can be obtained from subsequent measurements on the given physical system.

Recall that an instrument $\mathcal{N}$ is represented by a collection of observables $\mathcal{N}(A)[Y]$, defined in the following implicit fashion. For any particular A and Y, the expectation of the outcome of measuring the observable $\mathcal{N}(A)[Y]$ on a system in state ρ, is the same as the expectation of a function of the joint outcomes of *first* applying the *instrument* to a system in state ρ and *next* measuring the observable Y on the posterior state: namely, take the product of the indicator variable that the outcome of the instrument is in A, with the outcome of the subsequent measurement of Y. This collection of observables determines uniquely the probability distribution $\pi(\mathrm{d}x;\,\rho,\,\mathcal{N})$ of the outcome of applying the instrument $\mathcal{N}$ to the state ρ, and the posterior state $\sigma(x;\,\rho,\,\mathcal{N})$ given that the outcome is x. They are related to the $\mathcal{N}(A)[Y]$ by the equality (which we just expressed in words)

$$\mathrm{tr}\{\rho\mathcal{N}(A)[Y]\} = \int_A \mathrm{tr}\{\sigma(x;\,\rho,\,\mathcal{N})Y\}\pi(\mathrm{d}x;\,\rho,\,\mathcal{N}).$$

In the sequel we will drop the name of the instrument in the notation for π and σ and, when considering a parameterised family of prior states, replace the prior state $\rho(\theta)$ by the parameter value θ: thus $\pi(\mathrm{d}x;\,\theta)$ denotes the probability distribution of the outcome, and $\sigma(x;\,\theta)$ denotes the posterior state.

Definition 1 (Exhaustive instruments) A quantum instrument $\mathcal{N}$ is *exhaustive* for a parameterised set $\boldsymbol{\rho} : \Theta \to \mathcal{S}(\mathcal{H})$ of states if for all θ in Θ and for $\pi(\cdot;\,\theta)$-almost all x, $\sigma(x;\,\theta)$ does not depend on θ.

Thus the posterior states obtained from exhaustive quantum instruments are completely determined by the result of the measurement and do not depend on θ.

A useful strong form of exhaustivity is defined as follows.

Definition 2 (Completely exhaustive instruments) A quantum instrument $\mathcal{N}$ is *completely exhaustive* if it is exhaustive for all parameterised sets of states.

Recall that any completely positive instrument—in other words, virtually any physically realisable instrument—has the form (8) of $\mathcal{N}(A)[Y]$, given by

$$\mathcal{N}(\mathrm{d}x)[Y] = \sum_i \mathrm{tr}\{W_i(x)^* Y W_i(x)\}\nu(\mathrm{d}x) \tag{24}$$

with posterior states

$$\sigma(x;\rho) = \frac{\sum_i W_i(x)^* \rho W_i(x)}{\sum_i \mathrm{tr}\{\rho W_i(x) W_i(x)^*\}}$$

and outcome distributed as

$$\pi(\mathrm{d}x;\rho) = \sum_i \mathrm{tr}\{\rho W_i(x) W_i(x)^*\}\nu(\mathrm{d}x).$$

The following Proposition (which is a slight generalisation of a result of Wiseman 1999) shows one way of constructing completely exhaustive completely positive quantum instruments.

Proposition 1 *Let the quantum instrument $\mathcal{N}$ be as above, with $W_i(x)$ of the form*

$$W_i(x) = |\psi_x\rangle\langle\phi_{i,x}|, \tag{25}$$

for some functions $(i,x) \mapsto \phi_{i,x}$ and $x \mapsto \psi_x$. Then $\mathcal{N}$ is completely exhaustive.

PROOF. By inspection we find that the posterior state is

$$\sigma(x;\rho) = \frac{\sum_i |\phi_{i,x}\rangle\langle\phi_{i,x}|}{\sum_i \langle\phi_{i,x}|\phi_{i,x}\rangle},$$

which does not depend on the prior state ρ. □

5.2. Quantum Sufficiency

Suppose the measurement $M' = M \circ T^{-1}$ is a coarsening of the measurement M. In this situation we say that M' is *(classically) sufficient* for M with respect to a family of states $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ on $\mathcal{H}$ if the mapping T is sufficient for the identity mapping on $(\mathcal{X}, \mathcal{A})$ with respect to the family $\{P(\cdot;\theta;M) : \theta \in \Theta\}$ of probability measures on $(\mathcal{X}, \mathcal{A})$ induced by M and $\boldsymbol{\rho}$ (that is, $P(\cdot;\theta;M) = \mathrm{tr}\{M(\cdot)\rho(\theta)\}$).

As a further step towards a definition of quantum sufficiency, we introduce a concept of inferential equivalence of parametric models of states.

Definition 3 (Inferential equivalence) Two parametric families of states $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ and $\boldsymbol{\sigma} = \{\sigma(\theta) : \theta \in \Theta\}$ on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ are said to be *inferentially equivalent* if for every measurement M on $\mathcal{H}$ there exists a measurement N on $\mathcal{K}$ such that for all $\theta \in \Theta$

$$\mathrm{tr}\{M(\cdot)\rho(\theta)\} = \mathrm{tr}\{N(\cdot)\sigma(\theta)\} \tag{26}$$

and vice versa. (Note that, implicitly, the outcome spaces of M and N are assumed to be identical.)

In other words, $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ are equivalent if and only if they give rise to the same class of possible classical models for inference on the unknown parameter.

Example 10 (Two identical spin-half particles vs. one coherent spin-one particle) Let $\rho = \{\rho(\theta) : \theta \in \Theta\}$ be a parametric family of coherent spin-1 states; see Section 2.1.4 above. Then the associated Hilbert space $\mathcal{H}$ is $\mathbb{C}^2 \otimes \mathbb{C}^2$. Recall that the state vectors of coherent spin-1 states lie in the subspace $\mathcal{K} = \mathbb{C}^2 \odot \mathbb{C}^2$ of $\mathbb{C}^2 \otimes \mathbb{C}^2$. Define the parametric family $\sigma = \{\sigma(\theta) : \theta \in \Theta\}$ by $\sigma(\theta) = \Pi_{\odot}\rho(\theta)\iota$, where $\Pi_{\odot}$ and ι are the orthogonal projection from $\mathbb{C}^2 \otimes \mathbb{C}^2$ to $\mathbb{C}^2 \odot \mathbb{C}^2$ and the inclusion of $\mathcal{K}$ in $\mathcal{H}$, respectively. Given a measurement M on $\mathcal{H}$, we can define a measurement N on $\mathcal{K}$ by $N(\cdot) = \Pi_{\odot}M(\cdot)\iota$. Similarly, given a measurement N on $\mathcal{K}$, we can define a measurement M on $\mathcal{H}$ by $M(\cdot) = \iota N(\cdot)\Pi_{\odot}$. It is simple to verify that (26) is satisfied, and so ρ and σ are inferentially equivalent. □

Remark 1 It is of interest to find characterisations of inferential equivalence. This is a nontrivial problem, even in the case where the Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ are the same.

Next, let $\mathcal{N}$ denote an instrument on a Hilbert space $\mathcal{H}$ and with outcome space $(\mathcal{X}, \mathcal{A})$ and let $\mathcal{N}' = N \circ T^{-1}$ be a coarsening of $\mathcal{N}$ with outcome space $(\mathcal{Y}, \mathcal{B})$, generated by a mapping T from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$. According to (9) in Section 2.3, the posterior states for the two instruments are related by

$$\sigma(t; \theta, \mathcal{N}') = \int_{T^{-1}(t)} \sigma(x; \theta, \mathcal{N})\pi(\mathrm{d}x|t; \theta, \mathcal{N}),$$

where $\pi(\mathrm{d}x|t; \theta, \mathcal{N})$ is the conditional distribution of x given $T(x) = t$ computed from $\pi(\mathrm{d}x; \theta, \mathcal{N})$.

Definition 4 (Quantum sufficiency of instruments) Let $\mathcal{N}'$ be a coarsening of an instrument $\mathcal{N}$ by $T : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$. Then $\mathcal{N}'$ is said to be *quantum sufficient* with respect to a family of states $\{\rho(\theta) : \theta \in \Theta\}$ if

(i) the measurement $M'(\cdot) = \mathcal{N}'(\cdot)[\mathbf{1}]$ is sufficient for the measurement $M(\cdot) = \mathcal{N}(\cdot)[\mathbf{1}]$, with respect to the family $\{\rho(\theta) : \theta \in \Theta\}$

(ii) for any $x \in \mathcal{X}$, the posterior families $\{\sigma(x; \theta, \mathcal{N}) : \theta \in \Theta\}$ and $\{\sigma(T(x); \theta, \mathcal{N}') : \theta \in \Theta\}$ are inferentially equivalent.

5.3. Exhaustivity, Sufficiency, Ancillarity and Separability

In the theory of classical statistical inference, many important concepts (such as sufficiency, ancillarity and cuts) can be expressed in terms of the decomposition by a measurable function $T : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ of each probability distribution on $(\mathcal{X}, \mathcal{A})$ into the corresponding marginal distribution of $T(x)$ and the family of conditional distributions of x given $T(x)$. In quantum statistics there are analogous concepts based on the decomposition

$$\rho \mapsto (\pi(\cdot; \rho, \mathcal{N}), \sigma(\cdot; \rho, \mathcal{N})) \tag{27}$$

by a quantum instrument $\mathcal{N}$ of each state ρ into a measurement and a family of posterior states; see Section 2.3.

The classical concept of a cut encompasses those of sufficiency and ancillarity and is therefore more basic. A measurable function T is a *cut* for a set $\mathcal{P}$ of probability distributions on $\mathcal{X}$ if for all p_1 and p_2 in $\mathcal{P}$, the distribution on $\mathcal{X}$ obtained by combining the marginal distribution of $T(x)$

given by p_1 with the family of conditional distributions of x given $T(x)$ given by p_2 is also in $\mathcal{P}$; see, e.g. p. 38 of Barndorff-Nielsen and Cox 1994. Recent results on cuts for exponential models can be found in (Barndorff-Nielsen and Koudou, 1995), which also gives references to the useful role which cuts have played in graphical models. A generalisation to *local cuts* has become important in econometrics (Christensen and Kiefer, 1994, 2000). Replacing the decomposition into marginal and conditional distributions in the definition of a cut by the decomposition (27) yields the following quantum analogue.

Definition 5 (Quantum cuts) A quantum instrument $\mathcal{N}$ is said to be a *quantum cut* for a family $\boldsymbol{\rho}$ of states if for all ρ_1 and ρ_2 in $\boldsymbol{\rho}$, there is a ρ_3 in $\boldsymbol{\rho}$ such that

$$\begin{aligned} \pi(\cdot;\rho_3,\mathcal{N}) &= \pi(\cdot;\rho_1,\mathcal{N}) \\ \sigma(\cdot;\rho_3,\mathcal{N}) &= \sigma(\cdot;\rho_2,\mathcal{N}). \end{aligned}$$

Thus, if $\mathcal{N}$ is a quantum cut for a family $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ with ρ a one-to-one function then Θ has the product form $\Theta = \Psi \times \Phi$ and furthermore $\sigma(\cdot;\rho(\theta),\mathcal{N})$ depends on θ only through ψ, and $\pi(\cdot;\rho(\theta),\mathcal{N})$ depends on θ only through ϕ.

Example 11 (Simple quantum cuts) Let $\{\Pi_{[x]} : x \in \mathcal{X}\}$ be a PProM on a finite-dimensional quantum system. Suppose that sets Ψ and Ψ are given, together with collections of functions (indexed by x in $\mathcal{X}$) $f_x : \Phi \to [0,1]$ and $M_x : \Psi \to \mathcal{S}(\mathcal{H})$ which satisfy

$$\begin{aligned} \sum_{x\in\mathcal{X}} f_x(\phi) &= 1 && \phi \in \Phi \\ M_x(\psi) &= \Pi_{[x]} M_x(\psi) \Pi_{[x]} && \psi \in \Psi. \end{aligned}$$

Then we can define a family of states $\{\rho(\psi,\phi) : \psi \in \Psi, \phi \in \Phi\}$ by

$$\rho(\psi,\phi) = \sum_{x\in\mathcal{X}} f_x(\phi) M_x(\psi) \qquad (\psi,\phi) \in \Psi \times \Phi.$$

As indicated in Example 6, $\{\Pi_{[x]} : x \in \mathcal{X}\}$ gives rise to a simple quantum instrument $\mathcal{N}$, defined by

$$\mathcal{N}(\{x\})[Y] = \Pi_{[x]} Y \Pi_{[x]}.$$

A straightforward calculation using the orthogonality of the projections $\Pi_{[x]}$ shows that

$$\begin{aligned} \sigma(x;\rho(\psi,\phi),\mathcal{N}) &= M_x(\psi) \\ \pi(x;\rho(\psi,\phi),\mathcal{N}) &= f_x(\phi), \end{aligned}$$

and so $\mathcal{N}$ is a quantum cut for ρ. □

Since a quantum instrument is exhaustive for a parameterised set $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ of states if the family $\sigma(\cdot;\rho(\theta),\mathcal{N})$ of posterior states does not depend on θ, exhaustive quantum instruments are quantum cuts of a special kind. They can be regarded as quantum analogues of sufficient statistics. At the other extreme are the quantum instruments for which the measurements $\pi(\cdot;\rho(\theta),\mathcal{N})$ do not depend on θ. These can be regarded as quantum analogues of ancillary statistics.

Unlike exhaustivity, the concept of quantum sufficiency involves not only a quantum instrument but also a coarsening. The definition of quantum sufficiency can be extended to the following version involving parameters of interest.

Definition 6 (Quantum sufficiency for interest parameters) Let $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ be a family of states and let $\psi : \Theta \to \Psi$ map Θ to the space Ψ of interest parameters. A coarsening $\mathcal{N}'$ of an instrument $\mathcal{N}$ by a mapping T is said to be *quantum sufficient* for ψ on $\boldsymbol{\rho}$ if

(i) the measurement $\mathcal{N}'(\cdot)[\mathbf{1}]$ is sufficient for $\mathcal{N}(\cdot)[\mathbf{1}]$ with respect to the family $\boldsymbol{\rho}$,
(ii) for all θ_1 and θ_2 with $\psi(\theta_1) = \psi(\theta_2)$ and for all x in $\mathcal{X}$, the sets $\sigma(x; \rho(\theta_1), \mathcal{N})$ and $\sigma(T(x); \rho(\theta_2), \mathcal{N}')$ of posterior states are inferentially equivalent.

Consideration of the likelihood function obtained by applying a measurement to a parameterised set of states suggest that the following weakening of the concept of inferential equivalence may be useful.

Definition 7 (Weak likelihood equivalence) Two parametric families of states $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ and $\boldsymbol{\sigma} = \{\sigma(\theta) : \theta \in \Theta\}$ on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ respectively are said to be *weakly likelihood equivalent* if for every measurement M on $\mathcal{H}$ there is a measurement N on $\mathcal{K}$ with the same outcome space, such that

$$\frac{\mathrm{tr}\{M(\mathrm{d}x)\rho(\theta)\}}{\mathrm{tr}\{M(\mathrm{d}x)\rho(\theta')\}} = \frac{\mathrm{tr}\{N(\mathrm{d}x)\sigma(\theta)\}}{\mathrm{tr}\{N(\mathrm{d}x)\sigma(\theta')\}} \qquad \theta, \theta' \in \Theta$$

(whenever these ratios are defined) and vice versa.

Thus the likelihood function of the statistical model obtained by applying M to $\boldsymbol{\rho}$ is equivalent to that obtained by applying N to $\boldsymbol{\sigma}$, for the same outcome of each instrument.

Consideration of the distribution of the likelihood ratio leads to the following definition.

Definition 8 (Strong likelihood equivalence) Two parametric families of states $\boldsymbol{\rho} = \{\rho(\theta) : \theta \in \Theta\}$ and $\boldsymbol{\sigma} = \{\sigma(\theta) : \theta \in \Theta\}$ on Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ respectively are said to be *strongly likelihood equivalent* if for every measurement M on $\mathcal{H}$ with outcome space $\mathcal{X}$ there is a measurement N on $\mathcal{K}$ with some outcome space $\mathcal{Y}$ such that the likelihood ratios

$$\frac{\mathrm{tr}\{M(\mathrm{d}x)\rho(\theta)\}}{\mathrm{tr}\{M(\mathrm{d}x)\rho(\theta')\}} \qquad \text{and} \qquad \frac{\mathrm{tr}\{\mathrm{N}(\mathrm{dy})\sigma(\theta)\}}{\mathrm{tr}\{\mathrm{N}(\mathrm{dy})\sigma(\theta')\}}$$

have the same distribution for all θ, θ' in Θ, and vice versa.

The precise connection between likelihood equivalence and inferential equivalence is not yet known but the following conjecture appears reasonable.

Conjecture. Two quantum models are strongly likelihood equivalent if and only if they are inferentially equivalent up to quantum randomisation.

6. Quantum and Classical Fisher Information

In Section 3 we showed how to express the Fisher information in the outcome of a measurement in terms of the quantum score. In this section we discuss quantum analogues of Fisher information and their relation to the classical concepts.

6.1. Definition and First Properties

Differentiating (11) with respect to θ, writing $\rho_{/\!\!/\theta/\theta}$ for the derivative of the symmetric logarithmic derivative $\rho_{/\!\!/\theta}$ of ρ, and using the defining equation (10) for $\rho_{/\!\!/\theta}$ and the fact that ρ and $\rho_{/\!\!/\theta}$ are self-adjoint, we obtain

$$\begin{aligned} 0 &= \Re\,\mathrm{tr}\{\rho_{/\theta}(\theta)\rho_{/\!\!/\theta}(\theta) + \rho(\theta)\rho_{/\!\!/\theta/\theta}(\theta)\} \\ &= \Re\,\mathrm{tr}\left\{\tfrac{1}{2}\Big(\rho(\theta)\rho_{/\!\!/\theta}(\theta) + \rho_{/\!\!/\theta}(\theta)\rho(\theta)\Big)\rho_{/\!\!/\theta}(\theta)\right\} + \Re\,\mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta/\theta}(\theta)\} \\ &= I(\theta) - \mathrm{tr}(\rho(\theta)J(\theta))\,, \end{aligned}$$

where

$$I(\theta) = \mathrm{tr}\left\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)^2\right\}$$

is the *expected* (or *Fisher*) *quantum information*, already mentioned in Sections 3 and 4, and

$$J(\theta) = -\rho_{/\!\!/\theta/\theta}(\theta)\,,$$

which we shall call the *observable quantum information*. Thus

$$I(\theta) = \mathrm{tr}\{\rho(\theta)J(\theta)\}\ ,$$

which is a quantum analogue of the classical relation $i(\theta) = \mathrm{E}_\theta[j(\theta)]$ between expected and observed information (where $j(\theta) = -l_{/\theta/\theta}(\theta)$). Note that $J(\theta)$ is an observable, just as $j(\theta)$ is a random variable.

Neither $I(\theta)$ nor $J(\theta)$ depends on the choice of measurement, whereas $i(\theta) = i(\theta; M)$ does depend on the measurement M.

For parametric quantum models of states of the form

$$\boldsymbol{\rho} : \theta \mapsto \rho_1(\theta) \otimes \cdots \otimes \rho_n(\theta)$$

(which model 'independent particles'), the associated expected quantum information satisfies

$$I_{\rho_1\otimes\cdots\otimes\rho_n}(\theta) = \sum_{i=1}^{n} I_{\rho_i}(\theta)\,,$$

which is analogous to the additivity property of Fisher information. In particular, for parametric quantum models of states of the form

$$\boldsymbol{\rho} : \theta \mapsto \rho(\theta) \otimes \cdots \otimes \rho(\theta) \tag{28}$$

(which model n 'independent and identical particles'), the associated expected quantum information I_n satisfies

$$I_n(\theta) = nI(\theta)\,, \tag{29}$$

where $I(\theta)$ denotes the expected quantum information for a single measurement of the same type.

In the case of a multivariate parameter θ, the *expected quantum information* matrix $I(\theta)$ is defined in terms of the quantum scores by

$$I(\theta)_{jk} = \tfrac{1}{2}\mathrm{tr}\left\{\rho_{/\!\!/\theta_j}(\theta)\rho(\theta)\rho_{/\!\!/\theta_k}(\theta) + \rho_{/\!\!/\theta_k}(\theta)\rho(\theta)\rho_{/\!\!/\theta_j}(\theta)\right\}\ . \tag{30}$$

6.2. Relation to Classical Expected Information

Suppose that θ is one-dimensional. There is an important relationship between expected quantum information $I(\theta)$ and classical expected information $i(\theta; M)$, due to Braunstein and Caves (1994), namely that for any measurement M with density m with respect to a σ-finite measure ν on $\mathcal{X}$,

$$i(\theta; M) \leq I(\theta)\,, \tag{31}$$

with equality if and only if, for ν-almost all x,

$$m(x)^{1/2}\rho_{/\!\!/\theta}(\theta)\rho(\theta)^{1/2} = r(x)m(x)^{1/2}\rho(\theta)^{1/2}\,, \tag{32}$$

for some real number $r(x)$. For a proof see Appendix B.

For each θ, there are measurements which attain the bound in the quantum information inequality (31). For instance, we can choose M such that each $m(x)$ is a projection onto an eigenspace of the quantum score $\rho_{/\!\!/\theta}(\theta)$. Note that this attaining measurement may depend on θ.

Example 12 (Information for spin-half) Consider a spin-half particle in the pure state $\rho = \rho(\eta, \theta) = |\psi(\eta, \theta)\rangle\langle\psi(\eta, \theta)|$ given by

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

As we saw in Example 1 (where we wrote (η, ϑ) for (η, θ)), ρ can be written as $\rho = (\mathbf{1} + u_x\sigma_x + u_y\sigma_y + u_z\sigma_z)/2 = \frac{1}{2}(\mathbf{1} + \vec{u}\cdot\vec{\sigma})$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the three Pauli spin matrices and $\vec{u} = (u_x, u_y, u_z) = \vec{u}(\eta, \theta)$ is the point on the Poincaré sphere S^2 with polar coordinates (η, θ). Suppose that the colatitude η is known and exclude the degenerate cases $\eta = 0$ or $\eta = \pi$; the longitude θ is the unknown parameter.

Since all the $\rho(\theta)$ are pure, one can show that $\rho_{/\!\!/\theta}(\theta) = 2\rho_{/\theta}(\theta) = \vec{u}_{/\theta}(\theta)\cdot\vec{\sigma} = \sin(\eta)\,\vec{u}(\pi/2, \theta + \pi/2)\cdot\vec{\sigma}$. Using the properties of the Pauli matrices, one finds that the quantum information is

$$I(\theta) = \mathrm{tr}\{\rho(\theta)\rho_{/\!\!/\theta}(\theta)^2\} = \sin^2\eta.$$

Following Barndorff-Nielsen and Gill (2000), we now state a condition that a measurement must satisfy in order for it to achieve this information.

It follows from (32) that, for a pure spin-half state $\rho = |\psi\rangle\langle\psi|$, a necessary and sufficient condition for a measurement to achieve the information bound is: for ν-almost all x, $m(x)$ is proportional to a one-dimensional projector $|\xi(x)\rangle\langle\xi(x)|$ satisfying

$$\langle\xi|2\rangle\langle 2|a\rangle \;=\; r(x)\langle\xi|1\rangle\,,$$

where $r(x)$ is real, $|1\rangle = |\psi\rangle$, $|2\rangle = |\psi\rangle^{\perp}$ ($|\psi\rangle^{\perp}$ being a unit vector in $\mathbb{C}^2$ orthogonal to $|\psi\rangle$) and $|a\rangle = 2|\psi\rangle_{/\theta}$. It can be seen that geometrically this means that $|\xi(x)\rangle$ corresponds to a point on S^2 in the plane spanned by $\vec{u}(\theta)$ and $\vec{u}_{/\theta}(\theta)$.

If $\eta \neq \pi/2$, this is for each value of θ a different plane, and all these planes intersect in the origin only. Thus no single measurement M can satisfy $I(\theta) = i(\theta; M)$ for all θ. On the other hand, if $\eta = \pi/2$, so that the states $\rho(\theta)$ lie on a great circle in the Poincaré sphere, then the planes defined for each θ are all the same. In this case *any* measurement M with all components proportional to projector matrices for directions in the plane $\eta = \pi/2$ satisfies $I(\theta) = i(\theta; M)$ for all $\theta \in \Theta$. In particular, *any simple measurement in that plane* has this property.

More generally, a smooth one-parameter model of a spin-half pure state with everywhere positive quantum information admits a uniformly attaining measurement, i.e. such that $I(\theta) = i(\theta; M)$ for all $\theta \in \Theta$, if and only if the model is a great circle on the Poincaré sphere. This is actually a quantum exponential transformation model, see Example 9. $\square$

When the state ρ is strictly positive, and under further nondegeneracy conditions, essentially the only way to achieve the bound (31) is through measuring the quantum score. In the discussion below we first keep the value of θ fixed. Since any nonnegative self-adjoint matrix can be written as a sum of rank-one matrices (using its eigenvalue-eigenvector decomposition), it follows that any dominated measurement can be refined to a measurement for which each $m(x)$ is of rank 1, thus $m(x) = r(x)|\xi(x)\rangle\langle\xi(x)|$ for some real $r(x)$ and state-vector $|\xi(x)\rangle$, see the end of Section 2.2. If one measurement is the refinement of another, then the distributions of the outcomes are related in the same way. Therefore, under refinement of a measurement, Fisher expected information cannot decrease. Therefore if any measurement achieves (31), there is also a measurement with rank 1 components achieving the bound. Consider such a measurement. Suppose that $\rho > \mathbf{0}$ and that all the eigenvalues of $\rho_{/\!\!/\theta}$ are different. The condition $m(x)^{1/2}\rho_{/\!\!/\theta}\rho^{1/2} = r(x)m(x)^{1/2}\rho^{1/2}$ is then equivalent to $|\xi(x)\rangle\langle\xi(x)|\rho_{/\!\!/\theta} = r(x)|\xi(x)\rangle\langle\xi(x)|$, which states that $\xi(x)$ is an eigenvector of $\rho_{/\!\!/\theta}$. Since we must have $m(x)\mu(\mathrm{d}x) = \mathbf{1}$, it follows that all eigenvectors of $\rho_{/\!\!/\theta}$ occur in this way in components $m(x)$ of M. The measurement can therefore be reduced or coarsened (the opposite of refined) to a simple measurement of the quantum score, and the reduction (at the level of the outcome) is sufficient.

Suppose now the state $\rho(\theta)$ is strictly positive for all θ, and that the quantum score has distinct eigenvalues for at least one value of θ. Suppose a single measurement exists attaining (31) uniformly in θ. Any refinement of this measurement therefore also achieves the bound uniformly, in particular, the refinement to components which are all proportional to projectors onto orthogonal one-dimensional eigenspaces of the quantum score at the value of θ where the eigenvalues are distinct. Therefore the eigenvectors of the quantum score at this value of θ are eigenvectors at all other values of θ. Therefore there is a self-adjoint operator X with distinct eigenvalues such that $\rho_{/\!\!/\theta}(\theta) = f(X;\theta)$ for each θ. Fix θ_0 and let $F(X;\theta) = \int_{\theta_0}^{\theta} f(X;\theta)\mathrm{d}\theta$. Let $\rho_0 = \rho(\theta_0)$. If we consider the defining equation (10) as a differential equation for $\rho(\theta)$ given the quantum score, and with initial condition $\rho(\theta_0) = \rho_0$, we see that a solution is $\rho(\theta) = \exp\{\frac{1}{2}F(X;\theta)\}\rho_0\exp\{\frac{1}{2}F(X;\theta)\}$. Under smoothness conditions the solution is unique. Rewriting the form of this solution, we come to the following theorem:

Theorem 2 (Uniform attainability of quantum information bound) *Suppose that the state is everywhere positive, the quantum score has distinct eigenvalues for some value of θ, and is smooth. Suppose that a measurement M exists with $i(\theta; M) = I(\theta)$ for all θ, thus attaining the Braunstein–Caves information bound (31) uniformly in θ. Then there is an observable X such that a simple measurement of X also achieves the bound uniformly, and the model is of the form*

$$\rho(\theta) = c(\theta)\exp\{\tfrac{1}{2}F(X;\theta)\}\rho_0\exp\{\tfrac{1}{2}F(X;\theta)\} \tag{33}$$

for a function F, indexed by θ, of an observable X where $c(\theta) = 1/\mathrm{tr}\{\rho_0\exp(F(X;\theta))\}$, $\rho_{/\!\!/\theta}(\theta) = f(X;\theta) - \mathrm{tr}\{\rho(\theta)f(X;\theta)\}$, and $f(X;\theta) = F_{/\theta}(X;\theta)$. Conversely, for a model of this form, a measurement of X achieves the bound uniformly.

Remark 2 (Spin-half case) For spin-half, if the information is positive then the quantum score has distinct eigenvalues, since the outcome of a measurement of the quantum score always equals one of the eigenvalues, has mean zero, and positive variance. □

Theorem 3 (Uniform attainability of quantum Cramér–Rao bound) *Suppose the positivity and nondegeneracy conditions of the previous theorem are satisfied, and suppose that for the outcome of some measurement M a statistic t exists which is for all θ an unbiased estimator of θ achieving*

Helstrom's quantum Cramér–Rao bound (19), $\mathrm{Var}(t) = I(\theta)^{-1}$*. Then the model is actually a quantum exponential model of symmetric type (15),*

$$\rho(\theta) = c(\theta)\exp\{\tfrac{1}{2}\theta T\}\rho_0\exp\{\tfrac{1}{2}\theta T\} \tag{34}$$

for some observable T, and simple measurement of T is equivalent to the coarsening of M according to t.

PROOF. The coarsening of the measurement $M' = M \circ t^{-1}$ corresponding to t also achieves the quantum information bound (31) uniformly, $i(\theta; M') = I(\theta)$. Apply Theorem 2 to this measurement and we discover that the model is of the form (33), while (if necessary refining the measurement to have rank one components) t can be considered as a function of the outcome of a measurement of the observable X, and it achieves the classical Cramér–Rao bound for unbiased estimators of θ based on this outcome. Now the density of the outcome (with respect to counting measure on the eigenvalues of X) is found to be $c(\theta)\exp(F(x;\theta))\mathrm{tr}\{\rho_0\Pi_{[X=x]}\}$. Hence, up to addition of functions of θ or x alone, $F(x;\theta)$ is of the form $\theta t(x)$. □

Example 12 concerned pure spin-half models given by circles of constant latitude on the Poincaré sphere. Taking the product of n identical copies of such a model produces a spin-j model, with $j = n/2$, parameterised by a circle. It follows from the discussion at the end of Example 12, (29) and the additivity of Fisher information that if such a spin-j model is given by a great circle then there is a measurement M such that equality holds in (31).

The basic inequality (31) holds also when the dimension of θ is greater than one. In that case, the quantum information matrix $I(\theta)$ is defined in (30) and the Fisher information matrix $i(\theta; M)$ is defined by

$$i_{rs}(\theta; M) = \mathrm{E}_\theta[l_r(\theta)l_s(\theta)]\,,$$

where l_r denotes $l_{/\theta^r}$ etc. Then (31) holds in the sense that $I(\theta) - i(\theta; M)$ is positive semi-definite. The inequality is sharp in the sense that $I(\theta)$ is the smallest matrix dominating all $i(\theta; M)$. However it is typically not attainable, let alone uniformly attainable.

Theorem 2 can be generalised to the case of a vector parameter. This also leads to a generalisation of Theorem 3, which is the content of Corollary 1 below. First we give a lemma.

Lemma 1 *Let $\boldsymbol{\rho} : \Theta \to \mathcal{S}(\mathcal{H})$ be a twice differentiable parametric quantum model. Then*

$$(\rho_{/\!\!/j}\rho_{/\!\!/i} - \rho_{/\!\!/i}\rho_{/\!\!/j})\rho + \rho(\rho_{/\!\!/i}\rho_{/\!\!/j} - \rho_{/\!\!/j}\rho_{/\!\!/i}) = 2(\rho_{/\!\!/i/j} - \rho_{/\!\!/j/i}) \circ \rho,$$

where $\rho_{/\!\!/\theta} = (\rho_{/\!\!/1}, \ldots, \rho_{/\!\!/k})$ denotes the symmetric quantum score and $\circ$ denotes the Jordan product.

PROOF. By definition of $\rho_{/\!\!/\theta}$, we have

$$4\rho_{/i} = 2\left(\rho_{/\!\!/i}\rho + \rho\rho_{/\!\!/i}\right).$$

Differentiating this gives

$$\begin{aligned}4\rho_{/ij} &= 2\left(\rho_{/\!\!/i/j}\rho + \rho_{/\!\!/i}\rho_{/j} + \rho_j\rho_{/\!\!/i} + \rho\rho_{/\!\!/i/j}\right)\\ &= 2\left(\rho_{/\!\!/i/j}\rho + \rho\rho_{/\!\!/i/j}\right) + \rho_{/\!\!/i}\rho\rho_{/\!\!/j} + \rho_{/\!\!/i}\rho_{/\!\!/j}\rho + \rho\rho_{/\!\!/j}\rho_{/\!\!/i} + \rho_{/\!\!/j}\rho\rho_{/\!\!/i}.\end{aligned}$$

Since $\rho_{/ij} = \rho_{/ji}$, this leads to

$$\left(\rho_{/\!\!/j}\rho_{/\!\!/i} - \rho_{/\!\!/i}\rho_{/\!\!/j}\right)\rho + \rho\left(\rho_{/\!\!/i}\rho_{/\!\!/j} - \rho_{/\!\!/j}\rho_{/\!\!/i}\right) = 2\left\{\left(\rho_{/\!\!/i/j} - \rho_{/\!\!/j/i}\right)\rho + \rho\left(\rho_{/\!\!/i/j} - \rho_{/\!\!/j/i}\right)\right\}.$$

□

Theorem 4 *Let $\boldsymbol{\rho} : \Theta \to \mathcal{S}(\mathcal{H})$ be a twice differentiable parametric quantum model. If*

(i) there is a measurement M with $i(\theta; M) = I(\theta)$ for all θ,
(ii) $\rho(\theta) > \mathbf{0}$ for all θ,
(iii) Θ is simply connected

then, for any θ_0 in Θ, there are an observable X and a function F (possibly depending on θ_0) such that

$$\rho(\theta) = \exp\left\{\tfrac{1}{2}F(X;\theta)\right\}\rho(\theta_0)\exp\left\{\tfrac{1}{2}F(X;\theta)\right\}.$$

Proof. Since $i(\theta; M) = I(\theta)$, it follows from equation (32) and (iii) that there are real-valued functions $r_1, \ldots, r_{\dim\Theta}$ on $\mathcal{X} \times \Theta$ such that

$$m(x)\rho_{/\!\!/ i}(\theta) = r_i(x,\theta)m(x),$$

for all θ in Θ and ν-almost all x. Then

$$m(x)\rho_{/\!\!/ i}(\theta)\rho_{/\!\!/ j}(\theta') = r_i(x,\theta)r_j(x,\theta')m(x) = m(x)\rho_{/\!\!/ j}(\theta')\rho_{/\!\!/ i}(\theta),$$

for all θ, θ' in Θ and $1 \le i, j \le \dim\Theta$. Integration over $\mathcal{X}$ shows that $\rho_{/\!\!/ i}(\theta)$ and $\rho_{/\!\!/ j}(\theta')$ commute. By von Neumann's Theorem, there is an operator X and real-valued functions $f_1, \ldots, f_{\dim\Theta}$ on $\mathbb{R} \times \Theta$ such that

$$\rho_{/\!\!/ i}(\theta) = f_i(X;\theta). \tag{35}$$

Using condition (iii), and the fact that $\rho_{/\!\!/ i}$ and $\rho_{/\!\!/ j}$ commute, it follows from the Lemma that $\rho_{/\!\!/ i/j} = \rho_{/\!\!/ j/i}$. By condition (iv), (35) can be integrated to give a function F such that

$$\rho_{/\!\!/ i}(\theta) = F_{/i}(X;\theta).$$

The result follows by uniqueness of solutions of differential equations. □

Corollary 1 *If under the conditions of Theorem 4 there exists an unbiased estimator t of θ based on the measurement M achieving (19), then the model is a quantum exponential family of symmetric type (15) with commuting T_r.*

Versions of these results have been known for some time; see Young (1975), Fujiwara and Nagaoka (1995), Amari and Nagaoka (2000); compare especially our Corollary 1 to Amari and Nagaoka (2000, Theorem 7.6), and our Theorem 4 to parts (I) to (IV) of the subsequent outlined proof in Amari and Nagaoka (2000). Unfortunately the precise regularity conditions and detailed proofs seem to be available only in some earlier publications in Japanese. Note that we have obtained the same conclusions, by a different proof, in the spin-half pure state case, Example 12. This indicates that a more general result is possible without the hypothesis of positivity of the state.

The symmetric logarithmic derivative is not the unique quantum analogue of the classical statistical concept of score. Other analogues include the right, left and balanced derivatives obtained by suitable variants of (10). Each of these gives a quantum information inequality and a quantum Cramér–Rao bound analogous to (31) and (19). See Belavkin (1976). There is no general relationship between the various quantum information inequalities when the dimension of θ is greater than one.

In the next subsection we discuss the issue of asymptotic attainability of these and similar bounds.

6.3. Asymptotic Information Bounds

In classical statistics, the Cramér–Rao bound is attainable uniformly in the unknown parameter only under rather special circumstances. On the other hand, the restriction to unbiased estimators is hardly made in practice and indeed is difficult to defend. However, we have a richly developed asymptotic theory which states that in large samples certain estimators (e.g., the maximum likelihood estimator) are approximately unbiased and approximately normally distributed with variance attaining the Cramér–Rao bound. Moreover, no estimator can do better, in various precise mathematical senses (the Hájek–LeCam asymptotic local minimax theorem and convolution theorem, for instance). Recent work by Gill and Massar (2000), surveyed in Gill (2001a), makes a first attempt to carry over these ideas to quantum statistics. Similar results have been obtained, interestingly, with quite different methods, in a series of papers, by Young (1975), Fujiwara and Nagaoka (1995), Hayashi (1997), and Hayashi (1998). Another very recent approach, using large deviation theory rather than central limit theory, is given by Keyl and Werner (2001). The aim of Gill and Massar (2000) was to answer a question first posed by Peres and Wootters (1991): do joint measurements on a product of identical quantum systems contain more information about the common state of the subsystems, than separate measurements? The question was first answered—in the affirmative—in a rather specific form, by Massar and Popescu (1995): they considered for the most part just $n = 2$ copies of a spin-half pure state, in a Bayesian setting with a special loss function and prior distribution. Work of Barndorff-Nielsen and Gill (2000) showed that this advantage of joint over separate measurements disappears, for the spin-half pure state example, as $n \to \infty$.

The approach of Gill and Massar (2000) is firstly to delineate more precisely the class of attainable information matrices $i_n(\theta; M)$ based on arbitrary (or special classes) of measurements on the model (28) of n identical particles each in the same state $\rho(\theta)$. Next, using the van Trees inequality, a Bayesian version of the Cramér–Rao inequality, see Gill and Levit (1995), bounds on $i_n(\theta; M)$ are converted into bounds on the asymptotic scaled mean quadratic error matrix of regular estimators of θ. Thirdly, one constructs measurements and estimators which achieve those bounds asymptotically. The first step yields the following theorem.

Theorem 5 (Gill–Massar information bound) *In the model (28), one has*

$$\mathrm{tr}\{I(\theta)^{-1} i_n(\theta; M)/n\} \ \leq \ \dim(\mathcal{H}) - 1 \tag{36}$$

in any of the following cases: (i) $\dim(\theta) = 1$ *and* $\dim(\mathcal{H}) = 2$*, (ii)* ρ *is a pure state, (iii) the measurement* M *is separable.*

Case (i) follows from the earlier information inequality (31) from which follows, without any further conditions, $\mathrm{tr}\{I(\theta)^{-1} i_n(\theta; M)/n\} \leq \dim(\theta)$. The class of separable measurements, see Section 2.2, includes all *multilocal* instruments, i.e., instruments which are composed of a sequence of instruments acting on separate particles, see Section 2.3. Thus it is allowed that the measurement made on particle 2 depends on the outcome of the measurement on particle 1, and even that after these two measurements, yet another measurement, depending on the results so far, is made on the first particle in its new state, etc.

In the spin-half case the bound (36) is achievable in the sense that for any matrix K such that $\mathrm{tr}\{I(\theta)^{-1}K\} \leq 1$, there exists a measurement M on one particle, generally depending on θ, such that $i(\theta; M) = K$. The measurement is a randomised choice of several simple measurements of spin, one spin direction for each component of θ.

Application of the van Trees inequality gives the following asymptotic bound:

Theorem 6 (Asymptotic information bound) *In the model (28), let $V(\theta)$ denote the limiting scaled mean quadratic error matrix of a regular sequence of estimators $\widehat{\theta}_n$ based on a sequence of measurements M_n on n particles; i.e., $V^{ij}(\theta) = \lim_{n\to\infty} n\mathrm{E}_\theta\{(\widehat{\theta}^i_n - \theta^i)(\widehat{\theta}^j_n - \theta^j)\}$. Then V satisfies the inequality*

$$\mathrm{tr}\{I(\theta)^{-1}V(\theta)^{-1}\} \ \leq \ \dim(\mathcal{H}) - 1 \tag{37}$$

in any of the following cases: (i) $\dim(\theta) = 1$ *and* $\dim(\mathcal{H}) = 2$*, (ii) ρ is a pure state, (iii) the measurements M_n are separable.*

A *regular estimator sequence* is one for which the mean quadratic error matrices converge uniformly in θ to a continuous limit. It is also possible to give a version of the theorem in terms of convergence in distribution, Hájek-regularity and V the mean quadratic error matrix of the limiting distribution, rather than the limit of the mean quadratic error.

In the spin-half case, this bound is also asymptically achievable, in the sense that for any continuous matrix function $W(\theta)$ such that $\mathrm{tr}\{I(\theta)^{-1}W(\theta)^{-1}\} \leq 1$ there exists a sequence of separable measurements M_n with asymptotic scaled mean quadratic error matrix equal to W. This result is proved by consideration of a rather natural two-stage measurement procedure. Firstly, on a small (asymptotically vanishing) proportion of the particles, carry out arbitrary measurements allowing consistent estimation of θ, resulting in a preliminary estimate $\widetilde{\theta}$. Then on each of the remaining particles, carry out the measurement $\widetilde{M}$ (on each separate particle) which is optimal in the sense that $i(\widetilde{\theta}; \widetilde{M}) = K = W(\widetilde{\theta})^{-1}$. Estimate θ by maximum likelihood estimation, conditional on the value of $\widetilde{\theta}$, on the outcomes obtained in the second stage. For large n, since $\widetilde{\theta}$ will then be close to the true value of θ, the measurement $\widetilde{M}$ will have Fisher information $i(\theta; \widetilde{M})$ close to that of the 'optimal' measurement on one particle with Fisher information $i(\theta, M) = W(\theta)^{-1}$. By the usual properties of maximum likelihood estimators, it will therefore have scaled mean quadratic error close to $W(\theta)$. These measurements are not just separable, but multilocal, and within that class, adaptive and sequential with each new subsystem being measured only once.

In the spin-half case we have therefore a complete asymptotic efficiency theory in any of the three cases (i) a one-dimensional parameter, (ii) a pure state, (iii) separable measurements. By 'complete' we mean that it is precisely known what is the set of all attainable limiting scaled mean quadratic error matrices. This collection is described in terms of the quantum information matrix for one particle. What is interesting is that when none of these three conditions hold, greater asymptotic precision is possible. For instance, Gill and Massar (2000) exhibit a measurement of two spin-half particles which, for a completely unknown mixed state (a three-parameter model), has about 50% larger total Fisher information (for certain parameter values) than any separable measurement on two particles. Therefore if one has a large number n of particles, one has about 25% better precision when using the maximum likelihood estimator applied to the outcomes of this measurement on $n/2$ pairs of particles, than any separable measurement whatsoever on all n. It is not known whether taking triples, quadruples, etc., allows even greater increases of precision. It would be valuable to delineate precisely the set all attainable Fisher information matrices when non-separable measurements are allowed on each number of particles.

A similar instance of this phenomenon was called *non-locality without entanglement* by Bennett et al. (1999a). One could say that though the n particles are not in an entangled state, one needs an 'entangled measurement', presumably brought about by bringing the particles into interaction with one another (unitary evolution starting from the product state) before measurement, in order to extract maximal information about their state. The word 'non-locality' refers to the possibility that the n particles could be widely separated and brought into interaction through other entangled particles; see Section 8 for further examples of this kind in the context of optimal information transmission and in teleportation.

7. Infinite Dimensional Space

So far our examples have concerned spin-half systems, for which the dimension of the Hilbert space $\mathcal{H}$ is 2, and occasionally spin-j systems (dimension $2j+1$). In this section we give a survey of an important infinite dimensional example. The finite dimensional cases led us to parametric quantum statistical models. If the system has an infinite-dimensional Hilbert space, non- and semi-parametric quantum statistical models make an entrance. So far, they have been little studied from the point of view of modern mathematical statistics, despite their significance in experimental quantum physics, especially quantum optics.

7.1. Harmonic Oscillator

In this subsection we summarise some useful basic theory, and in the next we consider a basic statistical problem.

The simple harmonic oscillator is the basic model for the motion of a quantum particle in a quadratic potential well on the real line. Precisely the same mathematical structure describes oscillations of a single mode of an electromagnetic field (a single frequency in one direction in space). A useful orthonormal basis in the latter situation is given by the state-vectors of the pure states representing zero, one, two, ... photons. We denote these state-vectors by $|0\rangle, |1\rangle, |2\rangle, \ldots$. This basis is called the *number basis*. For the simple harmonic oscillator, the pure state with state-vector $|m\rangle$ is a state of definite energy $1/2+m$ units, $m=0,1,2,\ldots$. A pure state with state-vector $|\psi\rangle=\sum c_m|m\rangle$, where $\sum|c_m|^2=1$, is a complex superposition of these states. A mixed state ρ is a probability mixture over pure states $|\psi\rangle\langle\psi|$ with state-vectors $|\psi\rangle$.

Some key operators in this context, together with their common names, are

$$\begin{aligned}
A^+|n\rangle &= \sqrt{n+1}\,|n+1\rangle && \text{Creation}\\
A^-|n\rangle &= \sqrt{n}\,|n-1\rangle && \text{Annihilation}\\
N|n\rangle &= n\,|n\rangle && \text{Number}\\
Q &= (A^-+A^+)/\sqrt{2} && \text{Position}\\
P &= \frac{1}{i}(A^--A^+)/\sqrt{2} && \text{Momentum}\\
X_\phi &= (\cos\phi)\,Q+(\sin\phi)\,P && \text{Quadrature at phase } \phi\,.
\end{aligned} \tag{38}$$

One should observe that

$$\begin{aligned}
N &= A^+A^- = A^-A^+ - \mathbf{1} = \frac{1}{2}(Q^2+P^2-\mathbf{1})\\
[Q,P] &= i\mathbf{1}\,.
\end{aligned} \tag{39}$$

In the simple quantum harmonic oscillator, the state of a particle evolves under the Hamiltonian $H=\frac{1}{2}(Q^2+P^2)=N+\frac{1}{2}\mathbf{1}$; thus the state-vector of a pure state satisfies $|\psi(t)\rangle=e^{-iHt}|\psi(0)\rangle$, and an arbitrary state evolves as $\rho(t)=e^{-iHt}\rho(0)e^{iHt}$. The operators Q and P correspond to the position (on the real line) and the momentum of the particle. Indeed, the spectral decompositions of these two operators yield the PProM's of measurements of position and momentum respectively. It turns out that for a complex number $z=re^{i\phi}$ and the corresponding operator (called the

Weyl operator) $W_z = \exp(irX_\phi)$, we have $e^{i\theta N} W_z e^{-i\theta N} = W_{e^{i\theta}z}$, or in terms of the operator X_ϕ, $e^{i\theta N} e^{itX_\phi} e^{-i\theta N} = e^{itX_{\phi+\theta}}$. These relations become especially powerful when we note a short cut to the computation of the probability distribution of the measurement of the PProM corresponding to an observable X: it is the probability distribution with characteristic function $\mathrm{tr}\{\rho e^{itX}\}$. Combining these facts, we see that the distribution of the outcome of a measurement of position Q on the particle at time t is the same as that of X_t at time 0. In particular, with $t = \pi/2$, measuring P at time 0 has the same distribution as measuring Q at time $\pi/2$. For future reference, define $F = e^{-i(\pi/2)N}$ and note the relation $FP = QF$.

We mention for later reference that the Weyl operators form a projective unitary representation of the translation group on the real plane, since these are unitary operators with $W_z W_{z'} = w(z, z')W_{z+z'}$ for a certain complex function w of modulus 1, cf. (23).

In order to derive the probability distributions of outcomes of measurements of the observables defined above, it is useful to consider a particular concrete representation of the abstract Hilbert space $\mathcal{H}$ as $L^2_{\mathbb{C}}(\mathbb{R})$, that is, the space of complex-valued, Borel measurable, square integrable functions on the real line. The basis vectors $|n\rangle$ will be represented by normalised Hermite polynomials times the square root of the normal density with mean zero and variance half. The observables Q and P become rather easy to describe in this representation. At the same time, algebraic results from the theory of representations of groups provide further relations between the observables X_ϕ, N, Q and P.

Let us define the Hermite polynomials $H_n(x)$, $n = 0, 1, 2 \ldots$, by

$$H_n(x) = e^{x^2}(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}. \tag{40}$$

It follows that $H_n(x)$ is an n'th order polynomial with leading term $(2x)^n$. These polynomials can also be defined starting from the simple polynomials $(2x)^n$, $n = 0, 1, 2, \ldots$ by Gram–Schmidt orthogonalisation with respect to the normal density with mean 0 and variance $1/2$, $n(x) = (1/\sqrt{\pi})\exp(-x^2)$. Now if X is normal with mean zero and variance half, then $\mathrm{E}(H_n(X)^2) = 2^n n!$. Normalising, we obtain the following orthonormal sequence u_n in the space $L^2_{\mathbb{C}}(\mathbb{R})$:

$$u_n(x) = \sqrt{\frac{n(x)}{2^n n!}} H_n(x). \tag{41}$$

This sequence is not only orthonormal but complete—it forms a basis of $L^2_{\mathbb{C}}(\mathbb{R})$. The functions u_n satisfy the following recursion relations

$$\begin{aligned} \sqrt{2}\, x u_n(x) &= \sqrt{n+1}\, u_{n+1}(x) + \sqrt{n}\, u_{n-1}(x) \\ \frac{d}{dx} u_n(x) &= \sqrt{2}\sqrt{n}\, u_{n-1}(x) - x u_n(x). \end{aligned}$$

This shows us that under the equivalence defined by $|n\rangle \longleftrightarrow u_n$, one has the following correspondences

$$\begin{aligned} Q = (A^- + A^+)/\sqrt{2} &\longleftrightarrow x \\ P = \frac{1}{i}(A^- - A^+)/\sqrt{2} &\longleftrightarrow \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \\ 2N + \mathbf{1} = Q^2 + P^2 &\longleftrightarrow \left(x^2 - \frac{\mathrm{d}^2}{\mathrm{d}x^2}\right), \end{aligned} \tag{42}$$

where, on the first line, by 'x' we mean the operator of multiplication of a function of x by x to obtain a new function. In this representation the operator Q has 'diagonal' form, corresponding to the PProM with element $\Pi(B)$, B a Borel set of the real line, being the operator 'multiply by the indicator function 1_B'. Thus for a pure state with state-vector $|\psi\rangle$ in $\mathcal{H}$ represented by the wave-function $x \mapsto \psi(x)$ in $L^2_{\mathbb{C}}(\mathbb{R})$, the probability that a measurement of Q takes a value in B is equal to $\|1_B\psi\|^2 = \int_B |\psi(x)|^2 \mathrm{d}x$, so that the outcome of the measurement has probability density $|\psi(x)|^2$. Moreover,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} u_n(x) \mathrm{d}x = (-i)^n u_n(t) . \tag{43}$$

By expanding an arbitrary wave function ψ as a series of coefficients times u_n, one sees from this that the operator $F = e^{-i(\pi/2)N} = (-i)^N$ is nothing else than the Fourier transform, and its adjoint F^* is the inverse Fourier transform. The relation $FP = QF$ between Q and P involving F tells us that the probability distribution of a measurement of momentum P on a particle in the pure state with state-vector $|\psi\rangle$ has density equal to the absolute value of the square of the Fourier transform of the wave function $\psi(x)$. Measurement of Q is further studied in Example 14in Appendix A.1.

More generally, for the observable X_ϕ and considering mixed states instead of pure, from $e^{i\phi N} e^{itQ} e^{-i\phi N} = e^{itX_\phi}$ one may derive the following expression for the probability density of a measurement of X_ϕ on a system in state ρ:

$$p_\rho(x; \phi) = \sum_m \sum_{m'} \rho_{m,m'} e^{i(m-m')\phi} u_m(x) u_{m'}(x) , \tag{44}$$

where $\rho_{m,m'} = \langle m|\rho|m'\rangle$. The sense in which this double infinite sum converges is rather delicate; however, if only finitely many matrix elements $\rho_{m,m'}$ are non-zero, the formula makes sense as it stands.

7.2. Quantum Tomography

In this subsection we discuss a statistical problem, called for historical reasons *quantum tomography*, concerning the observables introduced in the previous subsection. Some key references are the book Leonhardt (1997) and the survey papers D'Ariano (1997a,b), though there has been much further progress since then. In its simplest form, the problem of quantum tomography is: given independent observations of measurements of the quadrature at phase ϕ, X_ϕ, with ϕ drawn repeatedly at random from the uniform distribution on $[0, 2\pi]$, reconstruct the state ρ. In statistical terms, we wish to do nonparametric estimation of ρ from n independent and identically distributed observations (ϕ_i, x_i), with ϕ_i as just described and x_i from the density (44) with $\phi = \phi_i$. In quantum optics, measuring a single mode of an electromagnetic field in what is called a quantum homodyne experiment, this would be the appropriate model with perfect photodetectors. In practice, independent Gaussian noise should be added.

Recalling that the probability density of a measurement of X_ϕ has $\mathrm{tr}\{\rho e^{it(\cos\phi Q + \sin\phi P)}\}$ as its characteristic function, we note that *if* Q and P were actually commuting operators (they are not!) then the joint characteristic function of a measurement of the two simultaneously would have been the function $\mathrm{tr}\{\rho e^{i(sQ+tP)}\}$ of the two variables (s, t).

Now the latter may not be the bivariate characteristic function of a joint probability density, but it is the characteristic function of a certain function called the Wigner function. This function $W_\rho(q, p)$ is known to characterise ρ. It is a real-valued function, integrating to 1 over the whole plane, but generally taking negative as well as positive values. The relation between the characteristic function of a measurement of X_ϕ and the characteristic function of the Wigner function which we have just described, shows that the probability density of a measurement of X_ϕ can be computed from the Wigner function by treating it as a joint probability density of two random variables $\widetilde{Q}$, $\widetilde{P}$ and computing from this density the marginal density of the linear combination $\cos\phi\,\widetilde{Q}+\sin\phi\,\widetilde{P}$. Now this computation is nothing else than a computation of the Radon transform of $W_\rho(q,p)$: its projection onto the line $(\cos\phi)q+(\sin\phi)p=0$. This transform is well known from computer-aided tomography, when for instance the data from which an X-ray image must be computed is the collection of one-dimensional images obtained by projecting onto all possible directions. Thus from the collection of all densities $p_\rho(x;\phi)$ of measurements of X_ϕ, one could in principle compute the Wigner function $W_\rho(q,p)$ by inverse Radon transform, from which one can compute other representations of ρ by further appropriate transformations. In particular, a double infinite integral over (p,q) of the product of the Wigner function with an appropriate kernel results in ρ in the 'position' representation, i.e., as the kernel of an integral transform mapping L^2 into L^2. Not all states can be so represented, but at least all can be approximated in this way. A further double infinite integral over (x,x') of another kernel results in ρ in the 'number' representation, i.e., the elements $\rho_{m,m'}$.

The basic idea of quantum tomography is to carry out this sequence of mathematical transformations on an empirical version of the density $p_\rho(x;\phi)$ obtained by some combination of smoothing and binning of the observations (ϕ_i,x_i). This theoretical possibility was discovered by Vogel and Risken (1989), and first carried out experimentally by M.G. Raymer and colleagues in path-breaking experiments in the early 1990's, see Smithey et al. (1993). Despite the enthusiasm with which the initial results were received, the method has a large number of drawbacks. To begin with, it depends on some choices of smoothing parameters and/or binning intervals, and later, during the succession of integral transforms, on truncations of infinite integrals among other numerical approximations. It has been discovered that these 'smoothings' tend to destroy precisely the interesting 'quantum' features of the functions being reconstructed. The final result suffers from both bias and variance, neither of which can be evaluated easily. Inverting the Radon transform is an ill-posed inverse problem and the whole procedure needs massive numbers of observations before it works reasonably well.

In the mid 1990's G.M. D'Ariano and his coworkers in Pavia have discovered a fascinating method to short-cut this approach, see D'Ariano (1997a,b). Using the fact that that the Weyl operators introduced above form an irreducible projective representation of the translation group on $\mathbb{R}^2$, they derived an elegant 'tomographic formula' expressing the mean of any operator A (not necessarily self-adjoint), i.e., $\mathrm{tr}(\rho A)$, as the integral of a function (depending on the choice of A) of x and ϕ, multiplied by $p_\rho(x;\phi)$, with respect to Lebesgue measure on $\mathbb{R}\times[0,2\pi]$. In particular, if we take the operator A to be $|m'\rangle\langle m|$ for given (m,m'), we have hereby expressed $\rho_{m,m'}$ as the mean value of a certain function, indexed by (m,m'), of the observations (ϕ_i,x_i), as long as the phases ϕ_i are chosen uniformly at random.

The key relation of their approach is the identity

$$A=\pi^{-1}\int_{\mathbb{C}}\mathrm{tr}(AW_z)W_{\overline{z}}\mathrm{d}z\,, \tag{45}$$

which can be derived (and generalised) with the theory of group representations. From this follows

$$\mathrm{tr}(\rho A)=\pi^{-1}\int_{\mathbb{C}}\mathrm{tr}(AW_z)\mathrm{tr}(\rho W_{\overline{z}})\mathrm{d}z\,. \tag{46}$$

The left hand side is the mean value of interest. The first 'trace' on the right hand side is a known function of the operator of interest A and the variable z. In the second 'trace' on the right hand side, after expressing $z = re^{i\phi}$ in polar coordinates, we recognise the characteristic function evaluated at the argument r of the probability density of our observations $p_\rho(x;\phi)$. Writing the characteristic function as the integral over x of e^{irx} times this density, transforming the integral over z into integrals over r and ϕ, and reordering the three resulting integrals, we can rewrite the right hand side as

$$\int_{x=-\infty}^{\infty}\int_{\phi=0}^{2\pi}\left[\int_{r=0}^{\infty} K_A(r,x,\phi)\mathrm{d}r\right] p_\rho(x;\phi)\mathrm{d}x\mathrm{d}\phi/(2\pi).$$

The innermost integral can sometimes be evaluated analytically, otherwise numerically; but in either case we have succeeded in our aim of rewriting means of operators of interest as means of known kernel functions of our observations. In the case $A = |m'\rangle\langle m|$, of interest for reconstructing $\rho_{m,m'}$, the kernel turns out to be bounded and hence we obtain unbiased estimators of the $\rho_{m,m'}$ with variance equal to $1/n$ times some bounded quantities.

Still this approach has its drawbacks. The required kernel function, in the case of reconstructing the density in the number representation, is highly oscillatory and even though everything is bounded, still huge numbers of observations are needed to get informative estimates. Also, the unbiased estimators constructed in this way are not unique and one may wonder whether better choices of kernels can be found. However, the approach does open a window of opportunity for further mathematical study of the mapping from $p_\rho(x;\phi)$ to $\rho_{m,m'}$ which could be a vital tool for developing the most recent approach, which we now outline briefly.

As we made clear, the statistical estimation problem seems related to the problems of nonparametric curve estimation, or more precisely, estimation of a parameter lying in an infinite dimensional space. Modern experience with such problems has developed an arsenal of methods, of which penalised and sieved likelihood, and nonparametric Bayesian methods, hold much promise as 'universal' approaches leading to optimal methods. In the present context, sieved maximum likelihood is very natural, since truncation of the Hilbert space in the number basis leads to finite dimensional parametric models which can in principle be tackled by maximum likelihood. One can hope that, from a study of the balance between truncation error (bias) and variance, it would be possible to derive data-driven methods to estimate ρ optimally with respect to a user-specified loss function. So far, only the initial steps in this research programme have been taken; in recent work Banaszek et al. (2000) and Paris et al. (2001) have shown that maximum likelihood estimation of the parameters in the density (44) is numerically feasible, after the number basis $\{|m\rangle : m = 0, 1, \ldots\}$ is truncated at (e.g.) $m = 15$ or $m = 20$. This means estimation of about 400 real parameters constrained to produce a density matrix. Numerical optimisation was used after a reparameterisation by writing $\rho = TT^*$ as the product of an upper-triangular matrix and its adjoint, so that only one constraint (trace 1) needs to be incorporated. We think that it is a major open problem to work out the asymptotic theory of this method, taking account of data-driven truncation, and possibly alleviating the problem of such a large parameter-space by using Bayesian methods. The method should be tuned to the estimation of various functionals of ρ of interest, and should provide standard errors or confidence intervals.

The quantum statistical model introduced above is that of optical homodyne measurements. There is also an elegant mathematical model for another experimental set-up called heterodyne measurement. In this case the measurement is a generalised measurement or OProM, and it can be *realised* by taking the product of the Hilbert space of the system of interest with another infinite dimensional system, in its ground state. Write Q', P' for the position and momentum operators on the ancillary system. It turns out that $P + P'$ and $Q - Q'$ commute, and therefore could in principle be measured simultaneously. A joint measurement of the two is a realisation of a heterodyne measurement. As an OProM it is invariant under the rotation group (corresponding to the phase changes ϕ of the homodyne measurement) and under a certain parametric model for the state, called the Gaussian or coherent state of the harmonic oscillator, possesses some decision-theoretic optimality properties because of this, see Holevo (1982). The pair now form a quantum transformation model in the sense of Section 4.2.

The field of quantum tomography is rapidly developing, with some of the latest (not yet published) results from the Pavia group of G.M. d'Ariano being *quantum holographic* methods to estimate not an unknown state, but an unknown transformation of a state (i.e., a completely positive instrument with trivial outcome space).

8. From Quantum Probability to Quantum Statistics

A recurring theme in this section is the relation between classical and quantum probability and statistics. This has been a matter of heated controversy ever since the discovery of quantum mechanics. It has mathematical, physical, and philosophical ingredients and much confusion, if not controversy, has been generated by problems of interdisciplinary communication between mathematicians, physicists, philosophers and more recently statisticians. Authorities from both physics and mathematics, perhaps starting with Feynman (1951), have promoted vigorously the standpoint that 'quantum probability' is something very different from 'classical probability'. Most recently, in two papers on Bell's inequality (which we discuss in Section 8.2) Accardi and Regoli (2000a,b), state "the real origin of the Bell's inequality is the assumption of the applicability of classical (Kolmogorovian) probability to quantum mechanics" which can only be interpreted as a categorical statement that classical probability is *not* applicable to quantum mechanics. Malley and Hornstein (1993) conclude from the perceived conflict between classical and quantum probability that 'quantum statistics' should be set apart from classical statistics.

We disagree. In our opinion, though fascinating mathematical facts and physical phenomena lie at the root of these statements, cultural preconceptions have also played a role. Statistical problems from quantum mechanics fall definitely in the framework of classical statistics and the claimed distinctions have retarded the adoption of statistical science in physics. The phenomenon of quantum entanglement in fact has far-reaching technological implications, which are easy to grasp in terms of classical probability; their development will surely involve statistics too.

In the first subsection we discuss, from a mathematical point of view, the distinction between classical and quantum probability. Next, we consider physical implications of the probabilistic predictions of quantum mechanics through the celebrated example of the Bell (1964) inequalities and the Aspect et al. (1982a,b) experiment. We appraise the 'classical versus quantum' question in the light of those implications. Finally we review a number of controversial issues in the foundations of quantum physics (locality, realism, the measurement problem) and sketch the basics of quantum teleportation, emphasizing that emerging quantum technology (entanglement-assisted communication, quantum computation, quantum holography and tomography of instruments) aims to capitalise on precisely those features of quantum mechanics which in the past have often been seen as paradoxical theoretical nuisances.

8.1. Classical versus Quantum Probability

Our stance is that the predictions which quantum mechanics makes of the real world are stochastic in nature. A quantum physical model of a particular phenomenon allows one to compute probabilities of all possible outcomes of all possible measurements of the quantum system. The word 'probability' means here: relative frequency in many independent repetitions. The word 'measurement' is meant in the broad sense of: macroscopic results of interactions of the quantum system under study with the outside world. These predictions depend on a summary of the state of the quantum system. The word 'state' might suggest some fundamental property of a particular collection of particles, but for our purposes all we need to understand under the word is: a convenient mathematical encapsulation of the information needed to make any such predictions. Some physicists argue that it is meaningless to talk of the state of a particular particle, one can only talk of the state of a large collection of particles prepared in identical circumstances; this is called a *statistical ensemble*. Others take the point of view that when one talks about the state of a particular quantum system one is really talking about a property of the mechanism which generated that system. Given that quantum mechanics predicts only probabilities, as far as real-world predictions are concerned the distinction between on the one hand a property of an ensemble of particles or of a procedure to prepare particles, and on the other hand a property of one particular particle, is a matter of semantics. However, if one would like to understand quantum mechanics by somehow finding a more classical (intuitive) physical theory in the background which would explain the observed phenomena, this becomes an important issue. It is also an issue for cosmology, when there is only one closed quantum system under study: the universe.

It follows from our standpoint that 'quantum statistics' is, for us, classical statistical inference about unknown parameters in models for data arising from measurements on a quantum system. However, just as in biostatistics, geostatistics, etc., etc., many of these statistical problems have a common structure and it pays to study the core ideas and common features in detail. As we have seen, this leads to the introduction of mathematical objects such as quantum score, quantum expected information, quantum exponential family, quantum transformation model, and so on; the names are deliberately chosen because of analogy and connections with the existing notions from classical statistics.

Already at the level of probability (i.e., before statistical considerations arise) one can see analogies between the mathematics of quantum states and observables on the one hand, and classical probability measures and random variables on the other. This analogy is very strong and indeed mathematically very fruitful (also very fruitful for mathematical physics). Note that collections of both random variables and operators can be endowed with algebraic structure (sums, products, ...). It is a fact that from an abstract point of view a basic structure in probability theory—a collection of random variables X on a countably generated probability space, together with their expectations $\int X \mathrm{d}P$ under a given probability measure P—can be represented by a (commuting) subset of the set of self-adjoint operators Q on a separable Hilbert space together with the expectations $\mathrm{tr}\{\rho Q\}$ computed using the trace rule under a given state ρ. Thus: a *basic* structure in classical probability theory is isomorphic to a *special case* of a basic structure in quantum probability. 'Quantum probability', or 'noncommutative probability theory' is the name of the branch of mathematics which studies the mathematical structure of states and observables in quantum mechanics. From this mathematical point of view, one may justly claim that classical probability is a special case of quantum probability. The claim does entail, however, a rather narrow view of classical probability. Moreover, many probabilists will feel that abandoning commutativity is throwing away the baby with the bathwater, since this broader mathematical structure has no analogue of the sample outcome ω, and hence no opportunity for a probabilist's beloved probabilistic arguments. We discuss Quantum Probability further in Section 9.1 under the heading of Quantum Stochastic Processes.

8.2. Bell, Aspect, et al.

We now discuss some physical predictions of quantum mechanics of a most striking 'nonclassical' nature. Many authors have taken this as a defect of classical probability theory and there have been proposals to abandon classical probability in favour of alternative theories (negative, complex or p-adic probabilities; nonmeasurable events; noncommutative probability; ...) in order to 'resolve the paradox'. However in our opinion, the phenomena are real and the defect, if any, lies in believing that quantum phenomena do not contradict classical *physical* thinking. This opinion is supported by the recent development of (potential) technology which acknowledges the extraordinary nature of the predictions and exploits the discovered phenomena (teleportation, entanglement-assisted communication, and so on). In other words, one should not try to explain away the strange features of quantum mechanics as some kind of defect of classical probabilistic thinking, but one should use classical probabilistic thinking to pinpoint these features.

Consider two spin-half particles, for which the customary state space is $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$. Let $|0\rangle$ and $|1\rangle$ denote the orthonormal basis of $\mathbb{C}^2$ corresponding to 'spin up' and 'spin down', thus two eigenvectors of the Pauli spin matrix σ_z. We write $|ij\rangle$ as an abbreviation for $|i\rangle \otimes |j\rangle$, defining four elements of an orthonormal basis of our $\mathcal{H}$.

For $\vec{u}$ in S^2, let $\sigma_{\vec{u}} = u_x\sigma_x + u_y\sigma_y + u_z\sigma_z$, the observable 'spin in the direction $\vec{u}$' for one spin-half particle. It has eigenvalues ± 1 and its eigenvectors are the state-vectors $\psi(\pm\vec{u})$ corresponding to the directions $\pm\vec{u}$ in S^2. The appropriate model for measurement of spin in direction $\vec{u}$ on the first particle and spin in the direction $\vec{v}$ on the second particle is a joint simple measurement of the two compatible observables $\sigma_{\vec{u}} \otimes \mathbf{1}$ and $\mathbf{1} \otimes \sigma_{\vec{v}}$ (see Example 6). The possible outcomes $\pm 1, \pm 1$ correspond to the one-dimensional subspaces spanned by the four orthogonal vectors $\psi(\pm\vec{u}) \otimes \psi(\pm\vec{v})$.

Now if the state of the system is a tensor product $\rho_1 \otimes \rho_2$ of separate states of each particle, then one can directly show that the outcomes for particle 1 and particle 2 are independent, and distributed as separate measurements on the separate particles, as one would hope. If the joint state is a mixture of product states, then the outcomes will be distributed as a mixture of independent outcomes. For an entangled state, the outcomes can be even more heavily dependent.

Consider the entangled pure state with state vector $\{|10\rangle - |01\rangle\}/\sqrt{2}$. This state is often called the *singlet* or *Bell* state. Straightforward calculations, see for instance Barndorff-Nielsen et al. (2002), show that for this state the two spin measurements have the following joint distribution: the marginal distribution of each spin measurement is Bernoulli($\frac{1}{2}$), the probability that the two outcomes are equal (both $+1$ or both -1) is $\frac{1}{2}(1 - \vec{u}\cdot\vec{v})$. In particular, if the two measurements are taken in the *same* direction, then the two outcomes are different with probability 1; in the opposite direction, the two outcomes are always the same; in orthogonal directions the probability of equality is $\frac{1}{2}$ so, taking account of the marginal distributions, the two outcomes are independent.

The singlet state is an appropriate description for the spins of two spin-half particles produced simultaneously in some nuclear scattering or decay processes where a total spin of 0 is conserved. The two particles have exactly opposite spin, which seems reasonable. The two particles are together in a pure state, which is also reasonable if the process involved was a Schrödinger evolution starting from a pure state. The model also exhibits a rotational invariance. These are all good reasons to expect the model to be not just a hypothetical possibility but a real possibility (and indeed, it is).

Fix a special choice of two possible different values of $\vec{u}$ and two possible different values of $\vec{v}$. Let us suppose that all four directions are in the same great circle on S^2 and let $\vec{u}_1$ and $\vec{u}_2$ be in the directions 0° and 120°, let $\vec{v}_1$ and $\vec{v}_2$ be in the directions 180° and 60°. Since $\cos(60°) = \frac{1}{2}$ we see that: when the directions are the pair (0°, 180°) then the probability the two spins are found to be equal is 1; but when the directions are any of the three pairs (0°, 60°) or (120°, 180°) or (120°, 60°)

the two spins are found to be equal with probability $\frac{1}{4}$. Is this surprising?

Consider an experiment where pairs of particles are generated in the singlet state, and then made to travel to two far-apart locations, at each of which spin is measured in one of the two directions just specified. Suppose the experiment is repeated many times, with random and independent choice of the two directions for measurement at each of the two locations. We have just computed the probabilities of all possible outcomes under each of the four possible combinations of directions.

Let us try to simulate the predicted statistics of the experiment using classical objects. To be very concrete, consider two people who try to simulate two spin-half particles. They start in a room together but then leave by different doors. Outside the room they are separately told a direction, $\vec{u}_1$ or $\vec{u}_2$ for person 1, $\vec{v}_1$ or $\vec{v}_2$ for person 2, and asked to choose an outcome '$+1$' or '-1'. They are not allowed to communicate any more once they have left the room. Moreover the directions will be chosen independently and randomly. The whole procedure will be repeated many many times and their aim is to simulate the quantum probabilities stated above. The two persons obviously will need randomisation in order to imitate the randomness of spin-half particles. We allow them to toss dice or coins, in any way they like, and to do this together in the room before leaving. They can simulate in this way any degree of dependence or independence they like. Let us call the outcome of their randomisation process ω. Their strategy will then be two pairs of functions of ω, with values ± 1, which determine the answers each person would give when confronted with each of his two directions on leaving the room, when the randomisation produces the outcome ω.

This whole set-up defines four Bernoulli ± 1-valued random variables, let us call them X_1, X_2, Y_1, Y_2; the X variables for person 1 and the Y variables for person 2. The four must be such that any pair X_i, Y_j has the same joint distribution as the result of measuring spins in the directions $\vec{u}_i$ and $\vec{v}_j$. Now it is easy to check that since these four variables are binary, $X_1 \neq Y_2$ and $Y_2 \neq X_2$ and $X_2 \neq Y_1$ implies $X_1 \neq Y_1$ (just fill in $+1$, -1, $+1$, -1 for X_1, Y_2, X_2, Y_1 in order; or alternatively -1, $+1$, -1, $+1$.) Conversely, therefore, $X_1 = Y_1$ implies $X_1 = Y_2$ or $Y_2 = X_2$ or $X_2 = Y_1$. Therefore we have

$$\mathrm{P}(X_1 = Y_1) \leq \mathrm{P}(X_1 = Y_2) + \mathrm{P}(Y_2 = X_2) + \mathrm{P}(X_2 = Y_1).$$

But the four probabilities we are trying to simulate are 1, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$ and it is not true that $1 \leq \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$. Therefore it is not possible to simulate with classical means (people or computers or other classical physical systems) the predicted outcomes of measurements of two spin-half particles!

The inequality we have just derived is due to Bell (1964) who contrasted it with the prediction of quantum mechanics in order to prove the failure, a priori, of any attempt through the introduction of *hidden variables* to explain the randomness of outcomes of measurements of quantum systems through 'mere statistical variation' in not directly observed and uncontrollable (hence hidden) properties of the quantum systems or measurement devices. He assumed that any physically meaningful hidden variables model would satisfy the physically reasonable property of *locality*, that is to say, the outcome of a measurement on one particle in one location should not depend on the measurement being carried out simultanously on the other particle in another distant location. Inspection of the argument we have given shows that Bell's inequality is not due to our slavish adherence to classical probability, but simply through the assumption that *the outcome of a measurement on one particle should not depend on* which *measurement is being made on the other particle*. This is reason enough for some authors, for instance Maudlin (1994), to conclude that Bell's argument shows that the predictions of quantum mechanics violate *locality*; he goes on to study the possible conflicts with relativity theory and concludes that there is no conflict in the sense that this phenomenon does not violate the requirements that cause and effect should not spread faster than the speed of light, and there is not a conflict with the basic relativistic (Minkowski) invariance property. Thus quantum mechanics lives in uneasy but peaceful coexistence with relativity theory.

All this would be purely academic were it not the case that the model we have just described truly is appropriate in certain physical situations and the predictions of quantum theory have been experimentally verified; first by Alain Aspect and his coworkers in a celebrated experiment (reported in Aspect et al. 1982a,b) in Orsay, Paris, where polarisation of pairs of entangled photons emitted from an excited caesium atom was measured with polarisation filters several metres apart; the orientation of the filters being fixed independently and randomly after the photons had been emitted from the source and before they arrived at the polarisation filter. (Polarisation of photons has a very similar mathematical description to spin of spin-half particles, except that all angles need to be halved: entangled photons have equal behaviour at polarisation filters oriented 90° to one another.) More recently, the experiment has been done on the glass fibre network of Swiss telecom with the two filters being 10 km apart on different shores of Lake Geneva.

Our conclusion is that quantum mechanics makes extraordinary physical predictions, predictions which are properly stated and interpreted in the language of classical probability. Technological implications of these predictions are only just beginning to be explored. One proposal is *entanglement-assisted communication*, see Bennett et al. (1999b, 2001); Holevo (2001b). Suppose A would like to send a message to B by encoding the message in the states of a sequence of spin-half particles transmitted one by one from A to B. At the receiving end B carries out measurements on the received particles on the basis of which he infers the message. Obviously the results will be random, especially if the communication channel suffers from noise, of classical or quantum nature. Using the theory of instruments one can describe mathematically all physically possible communication channels and all physically possible decoding (measurement) schemes, and compute analogously to classical information theory the maximum rate of transmission of information through the channel. Suppose now A and B allow themselves a further resource for communication. In between A and B a third person C is located, and he sends A and B simultaneously pairs of entangled spin half particles, in step with the transmission of particles from A to B. 'Obviously' there is no way these particles can be used to transmit information from A to B. They come from a different source altogether and are created in a fixed and known state. Yet it turns out that if A uses one part of the entangled pair in his encoding step with each particle he transmits, and B uses the other part of the pair in his decoding step, the rate of transmission can be *doubled*.

These extraordinary results show that it would be foolish to 'explain away' the phenomenon discovered by Bell by turning to some exotic probability theory (though many authors have done precisely this!). On the contrary, the mathematics—using classical probability—shows that strange things are going on and indeed it seems likely that one will be able to harness them in future technology.

8.3. Teleportation

As an example we show how the singlet state of a pair of spin-half particles, supposed to be in two distant locations, can be used to transmit a third spin-half state from one location to the other. This scheme was invented by Bennett et al. (1993) and experimentally carried out by A. Zeilinger's group in Innsbruck, see Bouwmeester et al. (1997). For a recent survey including references to the results of other experimental groups see Bouwmeester et al. (2001). The method illustrates how quantum technology (e.g., computation) will combine the basic ingredients of simple measurements, unitary evolution, and entanglement (product systems). The state being teleported is supposed to be completely unknown. This means that any attempt to measure it, and then teleport it by communicating in a classical way the results of measurement, cannot succeed, since the outcomes will be random, do not determine the initial state, and the initial state will have been destroyed by the measurement. The *no-cloning theorem* of Wootters and Zurek (1982), Dieks (1982) shows that there is no instrument which can transform a state ρ together with an ancillary quantum system into two identical copies $\rho \otimes \rho$.

Consider a single spin-half particle in the pure state with state-vector $\alpha|1\rangle + \beta|0\rangle$. It is brought into interaction with a pair of particles in the singlet state so that the whole system is in the pure state with state-vector, after multiplication of the tensor product, and up to a factor $1/\sqrt{2}$, $\alpha|110\rangle - \alpha|101\rangle + \beta|010\rangle - \beta|001\rangle$. The three particles are here written in the sequence: particle to be teleported, first entangled particle at the source location, second entangled particle at the destination location. Now we introduce the following four orthogonal state-vectors for the two particles at the source location, neglecting another constant factor $1/\sqrt{2}$, $\Phi_1 = |10\rangle - |01\rangle$, $\Phi_2 = |10\rangle + |01\rangle$, $\Psi_1 = |11\rangle + |00\rangle$, $\Psi_2 = |11\rangle - |00\rangle$, and we note that our three particles together are in a pure state with state-vector which may be written (up to yet another factor, $1/\sqrt{4}$) $\Psi_1 \otimes (\alpha|0\rangle - \beta|1\rangle) + \Psi_2 \otimes (\alpha|0\rangle + \beta|1\rangle) + \Phi_1 \otimes (-\alpha|1\rangle - \beta|0\rangle) + \Phi_2 \otimes (-\alpha|1\rangle + \beta|0\rangle)$. So far nothing has happened at all: we have simply rewritten the state-vector of the three particles as a superposition of four state-vectors, each lying in one of four orthogonal two-dimensional subspaces of $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$: namely the subspaces $\Phi_1 \otimes \mathbb{C}^2$, $\Phi_2 \otimes \mathbb{C}^2$, $\Psi_1 \otimes \mathbb{C}^2$ and $\Psi_2 \otimes \mathbb{C}^2$.

To these four subspaces corresponds a simple instrument. It only involves the two particles at the source location and hence may be carried out by the person at that location. He obtains one of four different outcomes, each with probability $\frac{1}{4}$, so he learns nothing about the particle to be teleported. However, conditional on the outcome of his measurement, the particle at the destination is in one of the four pure states with state-vectors $\alpha|0\rangle - \beta|1\rangle$, $\alpha|0\rangle + \beta|1\rangle$, $-\alpha|1\rangle - \beta|0\rangle$, $-\alpha|1\rangle + \beta|0\rangle$. The mixture with equal probabilities of these four states is the completely mixed state $\rho = \frac{1}{2}\mathbf{1}$, so nothing has happened at the destination: the state of the second part of the entangled pair still is in its original (marginal) state. But once the outcome of the measurement at the source is transmitted to the destination (two bits of information, transmitted by classical means), the receiver is able by means of one of four unitary transformations to transform the resulting pure state into the state with state-vector $\alpha|0\rangle + \beta|1\rangle$: teleportation is succesful. Neither source nor destination learn anything at all about the particle being transmitted by this procedure. If the state being teleported was a mixture, then decomposing it into pure components which are teleported independently and perfectly shows that the final destination state is the *same* mixture. In short, by transmitting two classical bits of information we are able to copy a point in the unit ball (specified by three real numbers) from A to B, without learning anything about the point at all in the process.

8.4. The Measurement Problem

We summarise here the problem raised by Schrödinger's cat, and survey briefly some responses. Consider a spin-half particle in the pure state with state-vector $\alpha|0\rangle + \beta|1\rangle$, where $|\alpha|^2 + |\beta|^2 = 1$. Suppose a measurement is made of the PProM with elements $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, resulting in the outcomes 0 and 1 with probabilities $|\alpha|^2$, $|\beta|^2$. Next to the measurement device is a cage containing a cat and a closed bottle of poison. If the outcome is 1, an apparatus automatically releases the poison and the cat dies. Otherwise, it lives. We suppose this whole system is enclosed in a large container and isolated from the rest of the universe.

Now the contents of that container are themselves just one large quantum system, and presumably it evolves unitarily under some Hamiltonian. If $\alpha = 0$, the final situation involves a dead cat. Let us denote its state-vector then by $|\text{dead}\rangle$. If $\beta = 0$ then the final state of the cat has state-vector $|\text{alive}\rangle$. So by linearity, in general the final state of the cat has state-vector $\alpha|\text{alive}\rangle + \beta|\text{dead}\rangle$. How would the cat experience being in this state?

When the container is opened and we look in, presumably a measurement does take place of the state of the cat, and at that moment (and only at that moment) it collapses into one of the two states with state-vectors $|\text{alive}\rangle$, $|\text{dead}\rangle$ with the probabilities $|\alpha|^2$, $|\beta|^2$. Recently, a number of experiments have been done which are purported to produce Schrödinger cats, in the sense of quantum superpositions of macroscopically distinct physical states of physical systems. For instance, Mooij et al. (1999) report on an experiment in which an electronic current involving of the order of a billion (10^9) electrons flows in a superposition of clockwise and anticlockwise directions around a supercooled alumuminium ring of a few micrometers in diameter (a thousand times larger than a typical molecular dimension). See Gill (2001b) for a discussion of this experiment and of the role of quantum statistics in confirming its success.

The situation is made more complicated when another person, known in the literature as Wigner's friend, is included in the system. He is in a room together with the container and at some point looks in the container. Only later does he report his findings to us.

This weird story accentuates some strange features of quantum mechanics. We told it as if 'the state' of a quantum system is something with physical reality, as it were, 'engraved' in the particles constituting the system. This idea leads us to suppose states exist which are very hard to imagine, and never observed in the real world. We see that the 'collapse of the state-vector', supposed to occur when a measurement takes place, seems to contradict the fact that measurement devices are physical systems themselves, and the device and the system being measured should evolve unitarily, not suddenly jump randomly from one state to another. We see that the dividing line between quantum system and the outside world is completely arbitrary, yet plays a central role in the theory (separating deterministic unitary evolution from random state-collapse).

Many different standpoints can be taken on these issues. The most extreme are those of the empiricist (or instrumentalist, or pragmatician) on the one hand, and the realist (who is actually an idealist) on the other. The empiricist does not believe in some kind of physical reality behind observed facts. He is interested only in making correct predictions about observable features of the world. For this person the only problem in our story is that the dividing line between quantum system and classical environment is somewhat arbitrary. If different descriptions lead to different prescriptions, there is a problem with the mathematical model. Below we present a simplified version of a consistency argument, which aims to show that there is no conflict between the two ingredients of quantum theory, and no inconsistency when the Heisenberg divide between quantum system and outside world may be placed at several different places.

Very similar considerations as those used in the consistency argument are also often used to argue that the von Neumann (random) collapse of the wave function can be *derived* from (deterministic) Schrödinger evolution. However we are inclined to believe that such claims are incomplete. If one believes that the state of things in the world is described by wave-functions, one still has a problem in relating wave-functions to physical properties of real objects. This problem is supposedly addressed by Everett's many worlds theory, van Fraassen's modal interpretation, and Griffiths' and Omnès' theory of consistent histories, among others. We find none of these attempts to make von Neumann redundant very convincing. However, the realist who wants the wave-function to be actually there in reality, and who believes that the true dynamics of physical systems is according to Schrödinger's equation alone, is forced in this direction. For cosmologists, wanting to model the whole universe without external observer, there seems to be a problem, since quantum randomness is a key part of modern theories of the origin of the universe.

The alternative for the realist is to extend or alter Schrödinger's dynamics in order to introduce a random element, which should make no difference to small quantum systems but should 'simulate' the von Neumann collapse, on big ones. Two fairly well explored variants of this idea are Bohm's hidden variables model, and the 'continuous spontaneous localisation' model of Ghirardi, Rimini and Weber. Most physicists are unhappy about these theories, since their claim to legitimacy is essentially that they *reproduce* unitary evolution and wave-function collapse in the two extreme situations where these should hold; 'in between' the physics is too difficult to make predictions, let alone test them by experiment. Thus the models do not seem to have new, testable consequences, while they include variables which determine the outcome of measurement, hence must be non-local.

Now we turn to the consistency argument, which aims to show that there is no contradiction between Schrödinger evolution and von Neumann collapse, in the sense that placing the dividing line between quantum system and outside world at different levels does not lead to different conclusions (at least, for an observer who is always in the outside world). This particular version was communicated to us by Franz Merkl.

Consider a spin-half particle which passes through the magnetic field of a Stern-Gerlach apparatus and then, if its spin is 'up', hits a photographic plate where a chain reaction produces a visible spot. If the spin is 'down' suppose the particle is lost. (This is a bit simpler than allowing the spin-down particle to hit the photographic plate at a different position: we have to model the interaction only in the spin-up case). We will call the photographic plate the detector. If the particle starts in the state $\alpha|0\rangle+\beta|1\rangle$, where where $|0\rangle$ and $|1\rangle$ represent spin-up and spin-down respectively, and the coefficients α and β satisfy $|\alpha|^2+|\beta|^2=1$, we get to see the spot with probability $|\alpha|^2$. Now the consistency problem arises because we could just as well have considered particle plus photographic plate as one large quantum system evolving jointly under some Hamiltonian for some length of time. If the detector started off in some pure state, then the final joint state of the joint system is another pure state, and no random jump to one of two possible final states has taken place. Let us however admit that the large system of the photographic plate involves many, many particles, and repetition of the experiment with the whole system in an identical pure state is physically meaningless to consider. At each repetition there are myriads of tiny differences. Therefore physically relevant predictions are only obtained when we use a mixed state as input for the macroscopic system. To make the mathematics even more simple, we will suppose that what varies from instance to instance is the length of time of the interaction. Let $|\psi\rangle$ be the state-vector of the detector, before the interaction starts. The joint system starts in the pure state with state-vector $(\alpha|0\rangle+\beta|1\rangle)\otimes|\psi\rangle$. Now the Hamiltonian of the interaction between particle and detector must be of the form $|0\rangle\langle 0|\otimes H$ where H acts on the huge Hilbert space of the detector, since there is a change to the detector if the particle starts in the spin-up state, but not at all if the particle starts in the spin-down state. Let the length of time of the interaction be τ. Then the final state of the joint system after the interaction is the pure state with state-vector $\alpha|0\rangle\otimes e^{-iH\tau/\hbar}|\psi\rangle+\beta|1\rangle\otimes|\psi\rangle$. The corresponding density-matrix can be written out, partitioned according to the first component of the joint system, as

$$\begin{pmatrix} |\alpha|^2 e^{-iH\tau/\hbar}|\psi\rangle\langle\psi|e^{iH\tau/\hbar} & \alpha\overline{\beta}e^{-iH\tau/\hbar}|\psi\rangle\langle\psi| \\ \overline{\alpha}\beta|\psi\rangle\langle\psi|e^{iH\tau/\hbar} & |\beta|^2|\psi\rangle\langle\psi| \end{pmatrix}$$

Now suppose we replace $H\tau$ by $H\tau+I\epsilon$ where I is the identity matrix. The idea here is that $H\tau$ must in some sense be large, since it produces a macroscopic change in a large quantum system. Thus this is a tiny perturbation of the interaction if ϵ is small, but on the other hand, since $\hbar$ is so tiny, $\epsilon/\hbar$ can still be very large. As we vary ϵ smoothly over some small interval, $\epsilon/\hbar$ varies smoothly over a huge range of values, and therefore the fractional part of $\epsilon/(2\pi\hbar)$ is close to uniformly distributed over the interval [0, 1]. Consequently, the factor $e^{-i\epsilon/\hbar}$ is close to uniformly distributed over the unit circle. Now after we have made this perturbation to the interaction, the density matrix of the joint state is

$$\begin{pmatrix} |\alpha|^2 e^{-iH\tau/\hbar}|\psi\rangle\langle\psi|e^{iH\tau/\hbar} & e^{-i\epsilon/\hbar}\alpha\overline{\beta}e^{-iH\tau/\hbar}|\psi\rangle\langle\psi| \\ e^{i\epsilon/\hbar}\overline{\alpha}\beta|\psi\rangle\langle\psi|e^{iH\tau/\hbar} & |\beta|^2|\psi\rangle\langle\psi| \end{pmatrix}.$$

On averaging over ϵ, the off-diagonal factors disappear and we find the density matrix

$$\begin{pmatrix} |\alpha|^2 e^{-iH\tau/\hbar}|\psi\rangle\langle\psi|e^{iH\tau/\hbar} & 0 \\ 0 & |\beta|^2|\psi\rangle\langle\psi| \end{pmatrix}.$$

This is the density matrix of the joint system which with probability $|\alpha|^2$ is in the pure state with state-vector $|0\rangle\otimes e^{-iH\tau/\hbar}|\psi\rangle$ and with probability $|\beta|^2$ is in the pure state with state-vector $|1\rangle\otimes|\psi\rangle$. In other words, either a spin-up particle and a detector which indicates a particle was detected, or a spin-down particle and a detector which indicates no particle was detected.

This argument is simple and one can criticise it in many ways. One would prefer to put the initial randomness into the many particles making up the detector, rather than into the interaction, and it should not have such a special form. But this is not a problem. Much more realistic models can be worked through which lead to the same qualitative conclusion: allowing variability in the initial conditions of the macroscopic measuring device, of a most innocuous kind, allows random phase factors such as $e^{-i\epsilon/\hbar}$ to wipe out off-diagonal terms in a large density matrix, so that all future predictions of the joint system are the same as if a random jump had occured during the initial interaction to one of two macroscopically distinct states.

In conclusion, it seems that as long as one is interested in using quantum mechanics only to predict what happens in a small part of the universe, and takes the randomness of quantum mechanics as intrinsic, not something which should be explained in a deterministic way, there are no logical inconsistencies in the theory. The state vector or state matrix of a quantum system should not be thought of as having an objective reality, somehow 'engraved' in the physical nature of a single instance of some quantum system, but is rather a characteristic of the preparation of the quantum system which, at least conceptually if not actually, could be repeated many times. Thus a statistical description goes in, and a statistical description comes out. The working quantum physicist even makes do without the von Neumann collapse of a quantum system, on measurement, since realistic quantum mechanical modelling of the quantum system under study *together* with the macroscopic measurement device allows one to introduce statistical variation in the initial state of the measurement device of the kind we have just described, and this leads irrevocably, it seems, to density matrices which are diagonal in the bases expressing macroscopically distinguishable states. In other words, unitary evolution alone, starting from the mixed initial state of quantum system plus measuring environment, is enough to determine the correct probability distribution over macroscopically distinguishable, thus 'real world', outcomes. The working quantum physicist is also well aware that the Hamiltonians he uses are only 'effective Hamiltonians' relative to some energy cut-off, which in turn corresponds to some approximation of a much larger state space by a smaller one. So the concerns of workers in the foundations of physics, worried about whether 'the state vector of the universe' evolves in a unitary, deterministic way, or a random, non-unitary way, could turn out in the long run to be as purely academic as those of medieval theologians trying to calculate how many angels could dance on the head of a pin, since sooner or later physicists will learn that quantum mechanics was itself only a limiting case of a better theory, as happened to Newtonian mechanics before. If we think about it carefully, we realise that the reality of basic concepts of classical physics is as illusory as that of basic concepts of modern physics.

9. Some Further Topics

9.1. Quantum stochastic processes

Since its inception in the early 1980's, through pioneering work of Hudson and Parthasarathy, quantum—or noncommutative—probability has grown into a mature and sophisticated mathematical field. The criticsm which we levelled at the philosphical standpoint of its protagonists in Section 8.1 does nothing to reduce the mathematical and physical results which have been achieved; see, for instance, Accardi et al. (1997). An excellent introduction to the field has been given by Biane (1995) and a more comprehensive account is available from the hand of Meyer (1993), see also Parthasarathy (1992). A new journal *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, now in its fourth year, is home to many of the more recent developments. Here we shall summarise briefly some aspects of quantum stochastic processes, under several subheadings.

Quantum optics Quantum optics is one of the currently most active and exciting fields of quantum physics, particularly from the viewpoint of the present paper. Laser cooling, on which we comment separately below, is, or may be viewed as, one of the areas in this field. Here we discuss briefly the Markov quantum (optical) master equation (MQME) and its quantum stochastic differential equation (QSDE) counterparts.

The Markovian quantum master equation provides an (approximate) description of a wide range of quantum system evolutions. The MQME is of the form

$$\dot{\rho}(t) = L(t)\rho(t),$$

where $L(t)$ is a linear operator. In order for this equation to have a solution such that $\rho(t)$ is a density operator for each t, $L(t)$ must be of the *Lindblad form* (Lindblad 1976)

$$L\rho = -\frac{i}{\hbar}[H,\rho] + \sum_k \Big(A_k \rho A_k^* - \frac{1}{2}\rho A_k^* A_k - \frac{1}{2} A_k^* A_k \rho \Big), \tag{47}$$

where H is some Hermitian operator and the A_k are (bounded) operators. To each such operator there exists a variety of QSDE's for a process $\psi(t)$ with values in $\mathscr{H}$ such that, writing $\hat{\rho}(t) = |\psi(t)\rangle\langle\psi(t)|/\langle\psi(t)|\psi(t)\rangle$, we have $\mathrm{E}[\hat{\rho}(t)] = \rho(t)$. See, for instance, Mølmer and Castin (1996), Wiseman (1996) and Gardiner and Zoller (2000, Chap. 5).

Interestingly, the same Markov quantum master equation has turned up in the Ghirardi-Rimini-Weber 'continuous spontaneous localisation' approach to the measurement problem, whereby unitary Schrödinger evolution is replaced by a stochastic differential equation, which is able to mimic, according to the circumstances, both purely unitary evolution of a closed quantum system, and the von Neumann collapse of the wave function of a quantum system interacting with a large (measuring) environment.

To illustrate how equation (47) can be numerically calculated by simulating many times a QSDE in what is called the quantum Monte Carlo approach, we consider the simplest case, when the index k just takes a single value and can therefore be omitted. Moreover, absorb the constant $\hbar$ into the Hamiltonian H. We show that the evolution is identical to the mean evolution of the following stochastic process for an unnormalised state vector ψ: the deterministic but non-Hamiltonian evolution

$$\dot{\psi} \;=\; -iH\psi - \frac{1}{2}A^*A\psi$$

interupted by *collapses*

$$\psi \;\rightarrow\; A\psi$$

with stochastic intensity

$$I \;=\; \|A\psi\|^2/\|\psi\|^2.$$

Introducing a counting process N with intensity I one can combine these equations into one QSDE of jump type,

$$\mathrm{d}\psi \;=\; (-iH\psi - \frac{1}{2}A^*A\psi)\mathrm{d}t + (A\psi - \psi)\mathrm{d}N.$$

Define $\widetilde{\rho} = \psi\psi^*$, the unnormalised random density matrix corresponding to the stochastic evolution, and $\widehat{\rho} = \widetilde{\rho}/\mathrm{tr}\,\widetilde{\rho}$. Note that $I = \mathrm{tr}(A\widetilde{\rho}A^*)/\mathrm{tr}\,\widetilde{\rho} = \mathrm{tr}(A\widehat{\rho}A^*)/\mathrm{tr}\,\widehat{\rho}$. Since $\mathrm{d}\widetilde{\rho} = \mathrm{d}\psi.\psi^* + \psi.\mathrm{d}\psi^*$ and $\dot{\psi}^* = i\psi^*H - \frac{1}{2}\psi^*A\psi$, the smooth part of the evolution can be rewritten as

$$\dot{\widetilde{\rho}} \;=\; -i[H,\widetilde{\rho}] - \frac{1}{2}(A^*A\widetilde{\rho} + \widetilde{\rho}A^*A).$$

Taking the trace, we find on the smooth part $\mathrm{d}(\mathrm{tr}\,\widetilde{\rho}) = -\mathrm{tr}(A\widetilde{\rho}A^*)$. Together, this yields

$$\begin{aligned}\frac{\mathrm{d}\widehat{\rho}}{\mathrm{d}t} \;&=\; \frac{1}{\mathrm{tr}\,\widetilde{\rho}}\frac{\mathrm{d}\widetilde{\rho}}{\mathrm{d}t} - \frac{\widetilde{\rho}}{(\mathrm{tr}\,\widetilde{\rho})^2}\frac{\mathrm{d}(\mathrm{tr}\,\widetilde{\rho})}{\mathrm{d}t}\\ &=\; -i[H,\widehat{\rho}] - \frac{1}{2}(A^*A\widehat{\rho} + \widehat{\rho}A^*A) + I\widehat{\rho}.\end{aligned}$$

For the jump part, define $N(t)$ to be the number of jumps in the time interval $(0,t]$. Then at a jump time we can write

$$\begin{aligned}\mathrm{d}\widehat{\rho} \;&=\; \Big(\frac{A\widehat{\rho}_-A^*}{\mathrm{tr}(A\widehat{\rho}_-A^*)} - \widehat{\rho}_-\Big)\mathrm{d}N\\ &=\; \Big(\frac{A\widehat{\rho}_-A^*}{I(t)} - \widehat{\rho}_-\Big)(\mathrm{d}N - I\mathrm{d}t) + (A\widehat{\rho}_-A^* - I\widehat{\rho}_-)\mathrm{d}t.\end{aligned}$$

Together this gives, at all time points,

$$\begin{aligned}\mathrm{d}\widehat{\rho} \;=\; &\Big(-i[H,\widehat{\rho}] - \frac{1}{2}(A^*A\widehat{\rho} + \widehat{\rho}A^*A) + A\widehat{\rho}A^*\Big)\mathrm{d}t\\ &+ \Big(\frac{A\widehat{\rho}_-A^*}{I(t)} - \widehat{\rho}_-\Big)(\mathrm{d}N - I\mathrm{d}t).\end{aligned}$$

Taking the expectation throughout, the martingale part (the second line) of this equation disappears, and $\widehat{\rho}$ in the first line is replaced by its expected value which we call ρ. The resulting nonstochastic differential equation for ρ is precisely (47). Moreover since $\widehat{\rho}$ was by construction a random density matrix (nonnegative, self-adjoint and trace one) we see that the solution ρ of (47), being the expected value of a density matrix, is also a density matrix; something which is not obvious from (47).

Example 13 (Quantum Monte Carlo for spin-half) Consider a two dimensional quantum system and choose a basis such that $H = E_1|1\rangle\langle 1| + E_2|2\rangle\langle 2|$, for real numbers E_1 and E_2. These are the two energy levels of the Hamiltonian. Suppose A is diagonal in this basis with $A|2\rangle = \alpha|1\rangle$ and $A|1\rangle = 0$ (the zero vector), where α is real. This is the model for the energy of a two-level atom which, on the spontaneous emission of a photon to its environment, can decay from its excited state to its ground state. Consider the evolution of an unnormalised state $\psi = c_1|1\rangle + c_2|2\rangle$, where c_1 and c_2 are complex functions of time. One discovers, since H and A^*A are simultaneously diagonalizable, that the smooth part of the evolution decouples as $\dot{c}_1 = (-iE_1 - \frac{1}{2}\alpha^2)c_1$, $\dot{c}_2 = (-iE_2)c_1$. Thus starting in state $|1\rangle$ or in state $|2\rangle$, we stay there, as long as no collapse occurs. If we are in state $|2\rangle$ collapse has intensity 0. However in state $|1\rangle$ there is a constant intensity α^2 of collapse to state $|1\rangle$. Thus starting in state $|1\rangle$, the QSDE predicts an exponential waiting time of collapse to $|2\rangle$ with rate α^2. The reader may like to compute the probability distribution of the time to collapse to state $|2\rangle$, starting from an arbitary pure state $\psi = \alpha|1\rangle + \beta|2\rangle$. □

As we remarked above, the same Lindblad equation can be represented as the mean evolution of a whole range of QSDE's, of jump type, diffusion type, and mixed type. Consider the same Lindblad equation as we were discussing above (no summation over k, drop $\hbar$). For an arbitrary real number μ define two matrices $D_\pm = (\mu\mathbf{1} \pm A)/\sqrt{2}$. Then the original Lindblad equation can be rewritten again in Lindblad form, with two different values of k, and the corresponding A_k being D_+ and D_-. This has a Quantum Monte Carlo representation of a smooth evolution $\dot{\psi} = (-iH - \frac{1}{2}D_+D_+{}^* - \frac{1}{2}D_-D_-{}^*)\psi$, interupted by collapses $\psi \to D_\pm\psi$ with intensities $\|D_\pm\psi\|^2/\|\psi\|^2$. The total intensity of jumps can be calculated as $\mu^2 + \|A\psi\|^2/\|\psi\|^2$. As $\mu \to \infty$ the rate of jumping increases without limit, but the relative change in the state at each jump becomes smaller and smaller. In the limit (after normalising suitably) one obtains a diffusion representation

$$\mathrm{d}\phi = \Big(-iH\phi + \frac{1}{2}\Big(\phi^*A\phi.A - \frac{1}{2}A^*A - \frac{1}{2}\phi^*A^*\phi.\phi^*A\phi\Big)\phi\Big)\mathrm{d}t + \frac{1}{2}\Big(2A - \phi^*A\phi - \phi^*A^*\phi\Big)\mathrm{d}W$$

where W is of course a standard Wiener process.

Laser cooling The paper by Mølmer and Castin (1996) on Monte Carlo techniques, for calculating expectation values for dissipative quantum systems, has been instrumental in particular in the context of laser cooling. Laser cooling is a topic of great current in interest in physics, both from the theoretical point of view and in terms of experimental advances opening up possibilities of studying many basic quantum phenomena, for instance Bose–Einstein condensation.

For a full understanding of the posssibility of subrecoil cooling, leading physicists were led to develop theoretical results that from the viewpoint of probability belong to renewal theory and add interesting new results and problems to that theory. For an introduction to this, see Barndorff-Nielsen and Benth (2001). A comprehensive account is given in Bardou, Bouchaud, Aspect, and Cohen-Tannoudji (2001). Barndorff-Nielsen, Benth, and Jensen (2000a,b) present some extensions to the setting of (classical) Markov processes.

Quantum infinite divisibility and Lévy processes Several types of quantum analogues of infinite divisibility and Lévy processes have recently been introduced. Two belong to free probability and are mentioned below. Infinitely divisible instruments and associated instrumental processes with independent increments are discussed in Holevo (2001a). See also Meyer (1993, Chap. 7), Barchielli and Paganoni (1996), and Albeverio, Rüdiger, and Wu (2001).

Free probability and random matrices The subject area of free probability evolves around the concept of free independence, also termed freeness. The latter was originally introduced by Voiculescu in the mid 1980's in a study of free-group von Neumann factors but was shortly afterwards realised to be naturally connected to the limiting properties of products of large and independent self-adjoint random matrices (of complex numbers). More specifically, suppose that $X_i^{(n)}$, $i = 1, \ldots, r$, are independent $n \times n$ random matrices, the entries in each of these matrices being also independent, and consider the mean values of the form

$$\mathrm{E}[\mathrm{tr}(X_{i_1}^{(n)} \ldots X_{i_p}^{(n)})]. \tag{48}$$

Under some mild regularity assumptions, for any given index set $i_1, \ldots, i_p$ and for $n \to \infty$, the quantity (48) will have a limiting value, and the collection of such mean values corresponds to a random limiting object. Freeness expresses how the independence of $X_1^{(n)}, \ldots, X_r^{(n)}$ is reflected in properties of that object. It is now possible to develop a theory of free infinite divisibility and free Lévy processes that to a large extent parallels that of infinite divisibility and Lévy processes in classical probability but also exhibit intriguing differences from the latter. There is, in particular, a one-to-one correspondence between the class of infinitely divisible laws in the classical sense and the class of the free infinitely divisible laws, with the 'free normal distribution' being the Wigner, or semicircle, law which has probability density

$$\pi^{-1}(1-x^2/2)^{1/2}.$$

This law was first derived by Wigner in the 1950's as the limiting law of the distribution of eigenvalues of a random Hermitian matrix $X^{(n)}$ with independent, complex Gaussian entries. Wigner's motivation for studying the eigenvalue distribution was based on the supposition that the local statistical behaviour of the energy levels of a sufficiently complex physical system is approximately simulated by that of the eigenvalues of a random matrix (Hamiltonian), see Wigner (1958) and Mehta (1967).

More detailed summaries of the mathematical connections indicated above are available in Biane (1998a,b) and Barndorff-Nielsen and Thorbjørnsen (2001). Furthermore, there are deep connections between the theory of random matrices and that of longest increasing subsequences, see for instance Deift (2000). We also wish to draw attention to a recent paper by Biane and Speicher (2001) which introduces a concept of free Fisher information.

General framework and continuous-time measurements The generic mathematical description of the measurement process embodied in formula (8) applies, in particular, to situations where a quantum system is observed continuously over a time interval $[0, T]$. For each time point $t \in [0, T]$, a representation such as in (8) is available for the data as available at that moment, but it is a highly non-trivial task, carried out by Loubenets (1999, 2000), Barndorff-Nielsen and Loubenets (2001) to mesh these representations together in an interpretable and canonical fashion. For simplicity, consider the case when the index i in (8) takes only one value and hence can be omitted. Often the outcome of a measurement of this type can be considered as the realisation of a cadlag stochastic process $x_0^T = \{x_t : 0 \le t \le T\}$ on $\mathbb{R}$ and the evolution of this and of the quantum system are determined by a probability measure ν on $D[0, T]$ and a collection of mappings $W_s^t(x_0^t)$, $0 \le s < t \le T$ from $\mathcal{X} = D[0, T]$ to $\mathbb{B}(\mathcal{H})$, satisfying the normalisation relations

$$\int_{D[0,T]} W_s^t(x_0^t)^* W_s^t(x_0^t)\nu(\mathrm{d}x_s^t | x_0^s) = I$$

and the cocycle conditions

$$W_s^t(x_0^t) = W_\tau^t(x_0^t) W_s^\tau(x_0^\tau)\,.$$

If the initial state of the quantum system in the Hilbert space $\mathcal{H}$ is a pure state ψ_0 then its evolutionary trajectory, conditional on x_0^T, is given by

$$\psi_t(x_0^t) = W_s^t(x_0^t)\psi_0\,.$$

Under suitable further conditions, the evolutions of x_t and ψ_t will be Markovian.

9.2. *Differential-geometric aspects*

In asymptotic parametric inference, differential geometry has proved to be an appropriate language for expressing various key concepts, see Barndorff-Nielsen and Cox (1994, Chaps. 5–7), Kass and

Vos (1997). Likewise, several concepts in quantum mechanics have differential-geometric interpretations. In particular, the quantum information $I(\theta)$ of a parametric quantum model is a Riemannian metric on the parameter space Θ, as is the Fisher information $i(\theta; M)$ obtained by a measurement M. There are many other Riemannian metrics of importance in quantum theory. A characterisation of a large class of them is given in Petz (1994). See also Petz and Sudár (1999). Any (complex) Riemannian metric on the space $\mathbb{SA}(\mathcal{H})$ of self-adjoint operators on a finite-dimensional $\mathcal{H}$ (and satisfying some mild conditions) yields an inequality analogous to Helstrom's quantum Cramér–Rao inequality (19). These inequalities and results on geometries obtained from suitable real-valued functions on $\Theta \times \Theta$ are given in Amari and Nagaoka (2000, Chap. 7). Some other differential-geometric aspects of quantum theory are considered in Brody and Hughston (2001).

9.3. Concluding Remarks

This paper has sketched some main features of quantum statistical inference, and more generally, quantum stochastic modelling. The basic concepts for our paper coincide with the basic concepts of quantum computation, quantum cryptography, quantum information theory, see Gruska (1999, 2001), Nielsen and Chuang (2000). We hope that many statisticians will venture into these areas too, as we are convinced that probabilistic modelling and statistical thinking will play major roles there, and should not be left purely to computer scientists or theoretical physicists.

Acknowledgements We gratefully acknowledge Mathematische Forschungsinstitut Oberwolfach for support through the Research in Pairs programme, and the European Science Foundation's programme on quantum information for supporting a working visit to the University of Pavia.

We have benefitted from conversations with many colleagues. We are particularly grateful to Elena Loubenets, Hans Maassen, Franz Merkl, Klaus Mølmer and Philip Stamp.

A. Mathematics of Quantum Instruments

Recall that an instrument $\mathcal{N}$ with outcomes x in the measurable space $(\mathcal{X}, \mathcal{A})$, is defined through a collection of observables $\mathcal{N}(A)[Y]$, for each $A \in \mathcal{A}$ and each bounded self-adjoint Y. With $\pi(\mathrm{d}x; \rho, \mathcal{N})$ denoting the probability distribution of the outcome of the measurement, and $\sigma(x; \rho, \mathcal{N})$ denoting the posterior state when the prior state is ρ and the outcome of the measurement is x, we have

$$\mathrm{tr}\{\rho\mathcal{N}(A)[Y]\} = \int_A \pi(\mathrm{d}x; \rho, \mathcal{N})\mathrm{tr}\{(\sigma(x; \rho, \mathcal{N})Y\}$$

Thus if one 'measures the instrument' on the state ρ, registers whether or not the outcome is in A, and subsequently measures the observable Y, the expected value of the outcome so obtained equals the expected value of the outcome of measuring directly the osbervable $\mathcal{N}(A)[Y]$.

A.1. Complete Positivity

The observables $\mathcal{N}(A)[Y]$ are sigma-additive in A, linear in Y, nonnegative in Y (map non-negative operators to non-negative operators), and normalised by $\mathcal{N}(\mathcal{X})[\mathbf{1}] = \mathbf{1}$. Any collection satisfying these constraints is called a *positive instrument*. Now given a positive instrument $\mathcal{N}$ defined on a Hilbert space $\mathcal{H}$, we can extend the instrument to the tensor product of this space with another Hilbert space $\mathcal{K}$ by defining $\mathcal{N}(A)[Y \otimes Z] = \mathcal{N}(A)[Y] \otimes Z$. This corresponds intuitively to measuring $\mathcal{N}$ on the first component of a quantum system in the product space, leaving the second component untouched. By linearity, once the extended instrument is defined on product observables like $Y \otimes Z$, it is defined on all observables of the product system. An instrument $\mathcal{N}$ is called *completely positive* if and only if every such extension (i.e., for any auxiliary system $\mathcal{K}$) remains positive. It turns out that one need only verify the positivity of the extensions for $\mathcal{K}$ of dimension $2, 3, \ldots, \dim(\mathcal{H}) + 1$.

Here is a classic example of an instrument which is positive, but not completely positive, hence is not physically realisable.

Example 14 (A positive, but not completely positive, instrument) Let the outcome space be trivial (consisting of a single element) so the instrument only transforms the incoming state, and does not generate any data. We therefore just specify an observable $\mathcal{N}[Y]$ for each observable Y: we define it by $\mathcal{N}[Y] = Y^\top$, the transpose of the observable Y. This corresponds to the outcome state $\sigma(\rho; \mathcal{N}) = \rho^\top$. Now take $\mathcal{K} = \mathcal{H}$, of finite dimension d, and define $|\psi\rangle = \frac{1}{d}\sum_i |i\rangle \otimes |i\rangle$ where the vectors $|i\rangle$ form an orthonormal basis of $\mathcal{H}$, take $\rho = |\psi\rangle\langle\psi|$. Let $\sigma = \rho^\top$ denote the corresponding output state. As a matrix operating on vectors, $\sigma(\sum_i c_i|i\rangle \otimes \sum_j d_j|j\rangle) = (\sum_i d_i|i\rangle \otimes \sum_j c_j|j\rangle)$. Thus in particular, σ maps $|i\rangle \otimes |j\rangle - |j\rangle \otimes |i\rangle$ to minus itself. Hence it has negative eigenvalues, and therefore cannot be a density matrix. □

Any dominated measurement M can be embedded into an instrument. The simplest way is by taking the posterior states to be $m(x)^{\frac{1}{2}}\rho m(x)^{\frac{1}{2}}/\mathrm{tr}(\rho m(x))$ for each outcome x having a positive density $\mathrm{tr}(\rho m(x))$ with respect to the same measure ν which dominates M. This corresponds to there being only one index i in (8), and $W(x) = m(x)^{\frac{1}{2}}$.

The next example illustrates the need to allow unbounded operators $W_i(x)$ in (8), even if the completely positive instrument in question is bounded.

Example 15 (Position measurement) As in Section 7.1 take as Hilbert space $\mathcal{H} = L^2_{\mathbb{C}}(\mathbb{R})$ and consider the PProM corresponding to the position observable Q. Thus the operator Q simply multiplies an L^2 function of x by the identity function $x \mapsto x$. The PProM has elements $M(B)$, for each Borel subset B of the real line, equal to the operator which multiplies an L^2 function by 1_B, the indicator function of the set B. In other words, $M(B)$ projects onto the subspace of functions which are zero outside B. The intuitively natural way to consider this measurement as part of an instrument would be to take the posterior state, given that the outcome is $x \in \mathbb{R}$, to be a delta-function at the point x. This is not an element of $\mathcal{H}$. However, one can easily imagine the following instrument $\mathcal{N}$: measure Q, and replace the quantum system by a new particle in the fixed state ρ_0, independently of the outcome x. (We reconsider the original instrument, later). By the physical interpretation of $\mathcal{N}(B)[Y]$, we must have, for any state ρ, that $\mathrm{tr}(\rho\mathcal{N}(B)[Y]) = \mathrm{tr}(\rho 1_B)\mathrm{tr}(\rho_0 Y)$. Suppose ρ_0 is the pure state with state vector $|\psi_0\rangle$. Then informally, in (8), one should have a single index i, dominating measure ν equal to Lebesgue measure, and $W(x) = |x\rangle\langle\psi_0|$ where the $|x\rangle$ stands for the delta-function at x, thus is not a particular member of $\mathcal{H}$, but is defined through the formula $\langle x|\psi\rangle = \psi(x)$. Thus $W(x)$ is the operator defined on the subspace of continuous L^2 functions ψ by $W(x)\psi = \psi(x)\psi_0$. It cannot be extended in a continuous way to all of L^2, and is therefore an unbounded operator. The instrument $\mathcal{N}$ can be written as $\mathcal{N}(\mathrm{d}x)[Y] = |\psi_0\rangle\langle\psi_0|\,\langle x|Y|x\rangle\,\mathrm{d}x$, or $\mathcal{N}(B)[Y] = |\psi_0\rangle\langle\psi_0|\langle 1_B|Y|1_B\rangle$, which is defined for all bounded operators Y and arbitrary Borel sets B.

Reconsider the instrument $\mathcal{N}'$ defined formally by $W(x) = |x\rangle\langle x|$. Formally, we should have $\mathcal{N}'(\mathrm{d}x)[Y] = |x\rangle\langle x|\langle x|Y|x\rangle\mathrm{d}x$ and thus $\mathcal{N}'(B)[Y] = \int_B |x\rangle\langle x|\langle x|Y|x\rangle\mathrm{d}x$. This formula is supposed to represent an observable, i.e., a possibly unbounded operator on $\mathcal{H}$. To find out what it does, we manipulate with delta-functions to find $\langle\phi|\mathcal{N}'(B)[Y]|\psi\rangle = \int_{\mathrm{R}} 1_B\overline{\phi}\psi\,\mathrm{d}\mu_Y$ where μ_Y is the finite measure on the real line defined by $\mu_Y(A) = \langle 1_A|Y|1_A\rangle$. Note that μ_Y is absolutely continuous with repect to Lebesgue measure ν. Thus $\mathcal{N}'(B)[Y]$ is defined on the subspace of L^2 functions, square integrable on B with respect to μ_Y, and on that subspace it acts by multiplying by the function $1_B \cdot \mathrm{d}\mu_Y/\mathrm{d}\nu$. The instrument $\mathcal{N}'$ is unbounded. It has an informal representation (8) involving objects W which cannot even be considered as unbounded operators, and there does not exist a posterior state for each outcome x of the instrument. There is a well-defined posterior state given the outcome lies in a set B of positive probability $\pi(B;\rho) = \mathrm{tr}(\rho 1_B)$. It is formally defined by $\sigma(B;\rho) = \int_B |x\rangle\langle x|\pi(\mathrm{d}x|B;\rho)$. □

A.2. Projection and Dilation of Measurements

Let $\Pi : \mathcal{H}' \to \mathcal{H}$ be the orthogonal projection of a Hilbert space $\mathcal{H}'$ onto a subspace $\mathcal{H}$. Then Π induces a map

$$\Pi^* : \mathrm{OProM}(\mathcal{X}, \mathcal{H}') \to \mathrm{OProM}(\mathcal{X}, \mathcal{H})$$

by

$$(\Pi^*(M))(A) = \Pi M(A)\Pi^* \qquad A \in \mathcal{A}\,. \tag{49}$$

In the physical literature, the OProM M is said to be a *dilation* or *extension* of $\Pi^*(M)$.

The following theorem shows that every OProM can be obtained from some PProM by the above construction: every generalised measurement can be dilated to a simple measurement.

Theorem 7 (Naimark 1940) *Given M in* OProM$(\mathcal{X}, \mathcal{H})$*, there is (i) a Hilbert space $\mathcal{H}'$ containing $\mathcal{H}$, (ii) a projection-valued probability measure M' in* PProM$(\mathcal{X}, \mathcal{H}')$*, such that*

$$\Pi^*\left(M'\right) = M$$

(in the sense of (49)), where $\Pi : \mathcal{H}' \to \mathcal{H}$ is the orthogonal projection.

The theorem of Naimark shows how to extend a generalised measurement to a simple measurement on a larger space. There is also an obvious way to consider a state on the smaller space as a state on the larger space, concentrating on the subspace. These two extensions together do not have the same statistical behaviour as the original pair of state and measurement. Adapting the proof of Naimark's theorem one can show how to extend an arbitrary state on the smaller space to a state on a larger space, in a way which matches the extension of the measurement, and together reproduces the statistics of the original set-up. This is taken care of by Holevo's theorem, Theorem 1 at the end of subsection 2.2.

B. The Braunstein–Caves Argument

A measurement M with density m with respect to a sigma-finite measure ν is given. Its outcome has density $p(x;\theta) = \mathrm{tr}\{\rho(\theta)m(x)\}$ with respect to ν. In the argument below, θ is also fixed. Define $\mathcal{X}_+ = \{x : p(x;\theta) > 0\}$ and $\mathcal{X}_0 = \{x : p(x;\theta) = 0\}$. Define $A = A(x) = m(x)^{\frac{1}{2}}\rho_{/\!/\theta}\rho^{\frac{1}{2}}$, $B = B(x) = m(x)^{\frac{1}{2}}\rho^{\frac{1}{2}}$, and $z = \mathrm{tr}\{A^*B\}$. Note that $p(x;\theta) = \mathrm{tr}\{B^*B\}$.

The proof of (31) given below consists of three inequality steps. The first will be an application of the trivial inequality $\Re(z)^2 \le |z|^2$ with equality if and only if $\Im(z) = 0$. The second will be an application of the Cauchy–Schwarz inequality $|\mathrm{tr}\{A^*B\}|^2 \le \mathrm{tr}\{A^*A\}\mathrm{tr}\{B^*B\}$ with equality if and only if A and B are linearly dependent over the complex numbers. The last step consists of replacing an integral of a nonnegative function over $\mathcal{X}_+$ by an integral over $\mathcal{X}$. Here they are:

$$\begin{aligned} i(\theta; M) &= \int_{\mathcal{X}_+} p(x;\theta)^{-1}(\Re\,\mathrm{tr}(\rho\rho_{/\!\!/\theta}m(x))^2\nu(\mathrm{d}x) \\ &\le \int_{\mathcal{X}_+} p(x;\theta)^{-1}|\mathrm{tr}(\rho\rho_{/\!\!/\theta}m(x))|^2\nu(\mathrm{d}x) \\ &= \int_{\mathcal{X}_+} \left|\mathrm{tr}\left(m(x)^{\frac{1}{2}}\rho^{\frac{1}{2}})^* \, (m(x)^{\frac{1}{2}}\rho_{/\!\!/\theta}\rho^{\frac{1}{2}} \right)\right|^2 (\mathrm{tr}(\rho m(x)))^{-1}\,\nu(\mathrm{d}x) \\ &\le \int_{\mathcal{X}_+} \mathrm{tr}(m(x)\rho_{/\!\!/\theta}\,\rho\rho_{/\!\!/\theta})\nu(\mathrm{d}x) \\ &\le \int_{\mathcal{X}} \mathrm{tr}(m(x)\rho_{/\!\!/\theta}\,\rho\rho_{/\!\!/\theta})\nu(\mathrm{d}x) \\ &= I(\theta). \end{aligned} \tag{50}$$

The necessary and sufficient conditions for equality at each of the three steps are therefore:

$$\begin{aligned} \Im(\mathrm{tr}\{A(x)^* B(x)\}) &= 0, \\ \alpha(x)A(x) + \beta(x)B(x) &= \mathbf{0}, \\ \int_{\mathcal{X}_0} \mathrm{tr}\{A(x)^* A(x)\}\nu(\mathrm{d}x) &= 0, \end{aligned}$$

where $\alpha(x)$ and $\beta(x)$ are arbitrary complex numbers, not both equal to zero, and the first two equalities are supposed to hold ν-almost everywhere where $p(x;\theta)$ is positive, while in the third equality $\mathcal{X}_0$ is precisely the set where $p(x;\theta)$ is zero.

Now if $A(x) = r(x)B(x)$ for real $r(x)$, for ν almost all x, then $A^*B = rB^*B$ and its trace is real. Hence the first and second conditions are satisfied. Moreover, we then also have $\mathrm{tr}\{A(x)A^*(x)\} = r(x)^2 p(x;\theta)$ so the third condition is also satisfied.

Conversely, suppose all three conditions are satisfied. Since $p(x;\theta) = \mathrm{tr}\{B(x)^*B(x)\}$, on $\mathcal{X}_+$ we must have B non-zero and hence α non-zero. So (still on $\mathcal{X}_+$) $A \propto B$ and the first condition implies that the proportionality constant must be real. The third condition implies that $\mathrm{tr}\{A(x)A(x)^*\}$ and hence $A(x)$ is almost everywhere zero where $p(x;\theta) = \mathrm{tr}\{B(x)^*B(x)\} = 0$, i.e., where $B(x) = \mathbf{0}$. So certainly one may write $A(x) = r(x)B(x)$ for some real $r(x)$ there, too.

In Braunstein and Caves' somewhat sketchy proof, it seems to be assumed that $p(x;\theta)$ is everywhere positive, hence only two inequality steps are involved. We note that the main ingredient of these proofs is the Cauchy–Schwarz inequality. This is also the main step in proving Helstrom's quantum Cramér–Rao bound, and of course in proving the classical Cramér–Rao bound.

References

Accardi, L., S. Kozyrev, and I. Volovich (1997). Dynamics of dissipative two-level systems in the stochastic approximation. *Phys. Rev. A* **56**, 2557–2562.

Accardi, L. and M. Regoli (2000a). Locality and Bell's inequality. Preprint, Volterra Institute, University of Rome II. `quant-ph/0007005`.

Accardi, L. and M. Regoli (2000b). Non-locality and quantum theory: new experimental evidence. Preprint, Volterra Institute, University of Rome II. `quant-ph/0007019`.

Albeverio, S., B. Rüdiger, and J.-L. Wu (2001). Analytic and probabilistic aspects of Lévy processes and fields in quantum theory. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes—Theory and Applications*, Boston. Birkhäuser.

Amari, S, I. and H. Nagaoka (2000). *Methods of Information Geometry*. Oxford: Oxford University Press.

Aspect, A., J. Dalibard, and G. Roger (1982a). Experimental realization of Einstein–Podolsky–Rosen–Bohm Gedankenexperiment: a new violation of Bell's inequalities. *Phys. Rev. Letters* **49**, 91–94.

Aspect, A., J. Dalibard, and G. Roger (1982b). Experimental test of Bell's inequalities using time-varying analysers. *Phys. Rev. Letters* **49**, 1804–1807.

Banaszek, K., G. D'Ariano, M. Paris, and M. Sacchi (2000). Maximum-likelihood estimation of the density matrix. *Phys. Rev. A* **61**, 010304(R).

Barchielli, A. and A. M. Paganoni (1996). A note on a formula of the Lévy–Khinchin type in quantum probability. *Nagoya Math. J.* **141**, 29–43.

Bardou, F., J. Bouchaud, A. Aspect, and C. Cohen-Tannoudji (2001). *Non-ergodic Cooling: Subrecoil Laser Cooling and Lévy Statistics*. Cambridge: Cambridge University Press. To appear.

Barndorff-Nielsen, O. and F. Benth (2001). Laser cooling and stochastics. In M. C. M. de Gunst, C. A. J. Klaassen, and A. W. van der Vaart (Eds.), *State of the Art in Probability and Statistics, Festschrift for W.R. van Zwet*, Lecture Notes–Monograph Series **36**, Hayward, Ca., pp. 50–71. Institute of Mathematical Statistics.

Barndorff-Nielsen, O., F. Benth, and J. L. Jensen (2000a). Light, atoms, and singularities. Research Report 2000-19, MaPhySto, University of Aarhus. (Submitted).

Barndorff-Nielsen, O., F. Benth, and J. L. Jensen (2000b). Markov jump processes with a singularity. *Ann. Appl. Prob* **32**, 779–799.

Barndorff-Nielsen, O. and E. Loubenets (2001). General framework for the behaviour of continuously observed open systems. Research Report 2001-??, MaPhySto, University of Aarhus.

Barndorff-Nielsen, O. and S. Thorbjørnsen (2001). Selfdecomposability and Lévy processes in free probability. *Bernoulli*. To appear.

Barndorff-Nielsen, O. E., P. Blæsild, J. L. Jensen, and B. Jørgensen (1982). Exponential transformation models. *Proc. Roy. Soc. London Ser. A* **379**, 41–65.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics*. London: Chapman and Hall.

Barndorff-Nielsen, O. E. and R. D. Gill (2000). Fisher information in quantum statistics. *J. Phys. A.: Math. Gen.* **33**, 4481–4490.

Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp (2001). Quantum Information. In B. Engquist and W. Schmid (Eds.), *Mathematics Unlimited—2001 and Beyond (Part I)*, Heidelberg, pp. 83–107. Springer.

Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp (2002). Quantum Stochastics. In Preparation.

Barndorff-Nielsen, O. E. and A. E. Koudou (1995). Cuts in natural exponential families. *Teor. Veroyatnost. i Primenen.* **2**, 361–372.

Belavkin, V. P. (1976). Generalized Heisenberg uncertainty relations, and efficient measurements in quantum systems. *Theoret. and Math. Phys.* **26**, 213–222.

Belavkin, V. P. (1994). Quantum diffusion, measurement and filtering I. *Theory Probab. Appl.* **38**, 573–585.

Belavkin, V. P. (2000). Quantum probabilities and paradoxes of the quantum century. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **3**, 577–610.

Belavkin, V. P. (2001). Quantum noise, bits, jumps: uncertainties, decoherence, measurements and filterings. *Progress in Quantum Electronics* **25**, 1–53.

Bell, J. S. (1964). On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200.

Bennett, C., G. Brassard, C. Crépeau, C. Jozsa, A. Peres, and W. Wootters (1993). Teleporting an unknown quantum state via dual classic and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899.

Bennett, C. H., D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters (1999a). Quantum nonlocality without entanglement. *Phys. Review A* **59**, 1070–1091.

Bennett, C. H., P. W. Shor, J. A. Smolin, and A. Thapliyal (1999b). Entanglement-assisted classical capacity of noisy quantum channels. *Phys. Rev. Lett.* **83**, 3081–3084.

Bennett, C. H., P. W. Shor, J. A. Smolin, and A. Thapliyal (2001). Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem. Technical report, AT&T Labs. `quant-ph/0106052`.

Biane, P. (1995). Calcul stochastique non-commutatif. In P. Bernard (Ed.), *Lectures on Probability Theory. Ecole d'Eté de Probabilités de Saint-Flour XXIII – 1993*, Lecture Notes in Mathematics **1608**, Heidelberg, pp. 1–96. Springer-Verlag.

Biane, P. (1998a). Free probability for probabilists. Preprint 40, MSRI.

Biane, P. (1998b). Processes with free increments. *Math. Zeitschrift* **227**, 143–174.

Biane, P. and R. Speicher (2001). Free diffusions, free entropy and free Fisher information. *Ann. Inst. H. Poincaré*. To appear.

Bouwmeester, D., J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger (1997). Experimental quantum teleportation. *Nature* **390**, 575–579.

Bouwmeester, D., J.-W. Pan, H. Weinfurter, and A. Zeilinger (2001). High-fidelity teleportation of independent qubits. *J. Modern Optics*. to appear; preprint `quant-ph/9910043`.

Brandt, S. and H. D. Dahmen (1995). *The Picture Book of Quantum Mechanics*. Heidelberg: Springer-Verlag.

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

Brody, D. C. and L. P. Hughston (2001). *The Geometry of Statistical Physics*. London/Singapore: Imperial College Press/World Scientific. To appear.

Christensen, B. J. and N. M. Kiefer (1994). Local cuts and separate inference. *Scand. J. Statistics* **21**, 389–401.

Christensen, B. J. and N. M. Kiefer (2000). Panel data, local cuts and orthogeodesic models. *Bernoulli* **6**, 667–678.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. London: Chapman and Hall.

D'Ariano, G. M. (1997a). Quantum estimation theory and optical detection. In T. Hakioğlu and A. S. Shumovsky (Eds.), *Quantum Optics and the Spectroscopy of Solids*, Amsterdam, pp. 135–174. Kluwer.

D'Ariano, G. M. (1997b). Measuring quantum states. In T. Hakioğlu and A. S. Shumovsky (Eds.), *Quantum Optics and the Spectroscopy of Solids*, Amsterdam, pp. 175–202. Kluwer.

Davies, E. B. (1976). *Quantum Theory of Open Systems*. London: Academic Press.

Davies, E. B. and J. T. Lewis (1970). An operational approach to quantum probability. *Comm. Math. Phys.* **17**, 239–260.

Deift, P. (2000). Integrable systems and combinatorial theory. *Notices AMS* **47**, 631–640.

Dieks, D. (1982). Communication by epr devices. *Phys. Lett. A* **92**, 271–272.

Feynman, R. P. (1951). The concept of probability in quantum mechanics. In *Proc. II Berkeley Symp. Math. Stat. and Prob.*, Berkeley, pp. 533–541. Univ. Calif. Press.

Fujiwara, A. and H. Nagaoka (1995). Quantum Fisher metric and estimation for pure state models. *Phys. Lett. A* **201**, 119–124.

Gardiner, C. and P. Zoller (2000). *Quantum Noise*. Berlin: Springer-Verlag. 2nd edition.

Gill, R. D. (2001a). Asymptotics in quantum statistics. In M. C. M. de Gunst, C. A. J. Klaassen, and A. W. van der Vaart (Eds.), *State of the Art in Probability and Statistics, Festschrift for W.R. van Zwet*, Lecture Notes–Monograph series **36**, Hayward, Ca., pp. 255–285. Institute of Mathematical Statistics.

Gill, R. D. (2001b). Teleportation into quantum statistics. *J. Korean Statist. Soc.*. in press.

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

Gill, R. D. and S. Massar (2000). State estimation for large ensembles. *Phys. Review A* **61**, 2312–2327.

Gilmore, R. (1994). *Alice in Quantum Land*. Wilmslow: Sigma Press.

Green, H. S. (2000). *Information Theory and Quantum Physics. Physical Processing for Understanding the Conscious Process*. Berlin: Springer.

Gruska, J. (1999). *Quantum Computation*. McGraw-Hill.

Gruska, J. (2001). Quantum Computing Challenges. In B. Engquist and W. Schmid (Eds.), *Mathematics Unlimited—2001 and Beyond (Part I)*, Heidelberg, pp. 529–563. Springer.

Hayashi, M. amd Matsumoto, K. (1998). Statistical model with a option for measurements and quantum mechanics. *RIMS koukyuroku* **1055**, 96–110.

Hayashi, M. (1997). A linear programming approach to attainable Cramér–Rao type bounds. In A. Hirota, A. Holevo, and C. Caves (Eds.), *Quantum Comunication, Computing and Measurement*, New York, pp. 99–108. Plenum.

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

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

Holevo, A. S. (2001a). Lévy processes and continuous quantum measurements. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes—Theory and Applications*, Boston. Birkhäuser.

Holevo, A. S. (2001b). On entanglement-assisted classical capacity. Preprint, Math. Inst., Russ. Acad. Sci. `quant-ph/0106075`.

Holevo, A. S. (2001c). *Statistical Structure of Quantum Theory*. Lecture Notes in Physics **m67**. Heidelberg: Springer-Verlag.

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

Kass, R. E. and P. W. Vos (1997). *Geometrical Foundations of Asymptotic Inference*. New York: Wiley.

Keyl, M. and R. Werner (2001). Estimating the spectrum of a density operator. Preprint, Inst. Math. Physik, T.U. Braunschweig. `quant-ph/0102027`.

Kraus, K. (1983). *States, Effects and Operations: Fundamental Notions of Quantum Theory*. Lecture Notes in Physics **190**. Berlin: Springer-Verlag.

Leonhardt, U. (1997). *Measuring the Quantum State of Light*. Cambridge: Cambridge University Press.

Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Comm. Math. Phys.* **48**, 119–130.

Loubenets, E. (1999). The quantum stochastic evolution of an open system under continuous in time nondemolition measurement. Research Report 1999-45, MaPhySto, University of Aarhus.

Loubenets, E. (2000). Quantum stochastic approach to the description of quantum measurements. *J. Phys. A.*. to appear; Research Report 2000-39, MaPhySto, University of Aarhus.

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

Massar, S. and S. Popescu (1995). Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263.

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

Mehta, M. (1967). *Random Matrices and the Statistical Theory of Energy Levels*. New York: Academic Press.

Meyer, P.-A. (1993). *Quantum Probability for Probabilists*. Lecture Notes in Mathematics **1538**. Berlin: Springer-Verlag.

Mølmer, K. and Y. Castin (1996). Monte Carlo wavefunctions. *Coherence and Quantum Optics* **7**, 193–202.

Mooij, J., T. Orlando, L. Levitov, L. Tian, C. van der Wal, and S. Lloyd (1999). Josephson persistent-current qubit. *Science* **285**, 1036–1039.

Naimark, M. A. (1940). Spectral functions of a symmetric operator. [in Russian with an English summary]. *Izv. Akad. Nauk SSSR, Ser. Mat.* **4**, 277–318.

Nielsen, M. and I. Chuang (2000). *Quantum Computation and Quantum Information*. New York: Cambridge University Press.

Ogawa, T. and H. Nagaoka (2000). Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Trans. Inf. Theory* **46**, 2428–2433.

Ozawa, M. (1985). Conditional probability and a posteriori states in quantum mechanics. *Publ. RIMS Kyoto Univ.* **21**, 279–295.

Paris, M., G. D'Ariano, and M. Sacchi (2001). Maximum-likelihood method in quantum estimation. Preprint, Dip. 'A. Volta', Univ. Pavia. `quant-ph/0101071`.

Parthasarathy, K. (1992). *An Introduction to Quantum Stochastic Calculus*. Basel: Birkhäuser.

Parthasarathy, K. (1999). Extremal decision rules in quantum hypothesis testing. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **2**, 557–568.

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

Peres, A. and W. K. Wootters (1991). Optimal detection of quantum information. *Phys. Rev. Lett.* **66**, 1119–1122.

Petz, D. (1994). Monotone metrics on matrix spaces. *Linear Algebra and its Applications* **244**, 81–96.

Petz, D. and C. Sudár (1999). Extending the Fisher metric toi density matrices. In O. E. Barndorff-Nielsen and E. V. Jensen (Eds.), *Geometry in Present Day Science*, Singapore, pp. 21–33. World Scientific.

Smithey, D., M. Beck, M. Raymer, and A. Faridani (1993). Measurement of the Wigner distribution and the density-matrix of a light mode using optical homodyne tomography—application to squeezed states and the vacuum. *Phys. Rev. Lett.* **70**, 1244–1247.

Stinespring, W. F. (1955). Positive functions on C^*-algebras. *Proc. Amer. Math. Soc.* **6**, 211–216.

Vogel, K. and H. Risken (1989). Determination of quasiprobability distributions in terms of probability-distributions for the rotated quadrature phase. *Phys. Rev. A* **40**, 2847–2849.

Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67**, 325–327.

Wiseman, H. M. (1996). Quantum trajectories and quantum measurement theory. *Quantum Semiclass. Opt.* **8**, 205–222.

Wiseman, H. M. (1999). Adaptive quantum measurements (summary). In *Miniproceedings: Workshop on Stochastics and Quantum Physics*, Miscellanea, 14, University of Aarhus. MaPhySto.

Wootters, W. and W. Zurek (1982). A single quantum cannot be cloned. *Nature* **299**, 802–803.

Young, T. Y. (1975). Asymptotically efficient approaches to quantum-mechanical parameter estimation. *Information Sciences* **9**, 25–42.