Lecture Notes on Quantum Statistics

Richard D. Gill
Part of book project with
O.E. Barndorff-Nielsen and P.E. Jupp

Version: 30 January 2000

Preface

These notes are meant to form the material for an introductory course on quantum statistics at the graduate level, aimed at mathematical statisticians and probabilists. No background in physics (quantum or otherwise) is required. They are still far from complete.

Quantum statistics as we mean it here is statistical inference based on data obained from measurement of a quantum system. The reader is probably aware that quantum physics makes stochastic predictions about reality. The actual outcome of an experiment involving measurements on some small number of elementary particles cannot be predicted. Rather, quantum mechanics allows one to compute the probability distribution of the outcomes. This probability distribution will depend on a specification of the system under study and on the chosen measurement apparatus. Often such a specification depends on parameters which are not known in advance and then the data could be used to gain information about them. These parameters could correspond to classical, macroscopic features of the system producing these elementary particles. Examples could be the orientation or position of apparatus producing a stream of photons in a quantum communication system, or properties of some distant star or other astronomical object so distant or weak that we can only detect a finite number of photons from this source during a finite observational period.

In the past, physical predictions made on quantum systems typically involved huge numbers of identical particles and focussed on their collective behaviour. The random nature of the outcome was submerged by the law of large numbers, and the aim was to compute expectated values and correlations. Thus quantum physics provided exact predictions of non-random, aggregate properties. Statistical questions might have been important but the underlying stochastic nature of the phenomenon under study at the level of individual particles did not play a role. Nowadays however experimentation and theory are focussing more and more on manipulating truly small quantum systems consisting of just one or a really small number of atoms, electrons, photons or whatever. Technology will surely follow. Theory and

speculation on such topics as quantum computing, quantum cryptography, and quantum communication channels, far outrun experiment and technology but these are also developing fast and will in the near future, we believe, supply real life statistical problems prototypical forms of which are studied in these notes.

Does the reader need a background in physics in order to study quantum statistics? We believe not. A beautiful feature of quantum mechanics is that it can be described abstractly on the basis of a fairly straightforward mathematical apparatus: the mathematics of linear operators on a complex Hilbert space (the so-called state-space) together with some elementary probability theory. Moreover, when we are concerned purely with 'discrete' properties of a quantum system such as spin of a finite number of electrons, polarization of a finite number of photons, the energy level (ground state or excited) of a finite array of atoms at very low temperature, and can ignore other aspects of the system under study, finite-dimensional Hilbert spaces suffice. Thus the mathematical model can be formulated in the language of finite dimensional complex vectors and matrices. Already in such a finite-dimensional set-up one can pose and try to solve important and non-trivial problems of quantum statistical inference. And these problems could turn up in quite diverse applications since the basic model could apply to different aspects of quite different systems. We do not yet know if the particular 'toy' problems we consider will turn up in practice in the coming years, but we are confident that what we have learnt from their study will be of value in future applications.

The notes concentrate on the natural statistical problems arising from the simplest possible quantum settings. Moreover we study whether various key notions in mathematical statistics—such as exponential families or transformation models—have extensions or analogues in quantum statistics. In particular the statistical notion of information, and information bounds based on the Fisher information matrix, have already turned out to be very useful in quantum statistical design and inference problems. As we have said, the results of measurement of a quantum system are random, and quantum theory tells us how to compute the resulting probability distributions. Consequently quantum statistical inference is 'ordinary' statistics applied to the kind of models which turn up in quantum mechanics.

Quantum theory places fundamental limits to the amount of information which can be obtained from a single quantum system. We have a very precise description of the class of all possible measurements on a given quantum system and this generates fundamental limits to the precision with which the state of the system can be determined through measurements on the system. Moroever one can distinguish subclasses of measurements which

might be more easy to implement in practice. Thus 'experimental design' problems with a clean mathematical description arise very naturally.

Despite the enormous success of quantum mechanics, its foundations are still the subject of much controversy. This could be either an appealing or an off-putting feature of this problem area for readers, depending on their personal inclinations. Fortunately we do not need to take sides concerning foundational aspects of quantum mechanics. We simply take the model of quantum mechanics as a description of empirical reality. However the reader who is interested in foundational aspects—which involve what we mean by randomness and determinism, and hence could be of interest to philosophically inclined probabilists and statisticians—can meet all the paradoxical features of quantum mechanics arising from wave-particle duality such as entanglement, non-locality, and the measurement problem, in the context of simple finite-dimensional models. The controversies concern the proper interpretation and consistency of this mathematical picture. We plan to include some introductory material on these matters but emphasize that it is not necessary to take sides or even become aquainted with these problems in order to solve concrete quantum statistical problems, just as quantum physicists manage very well to make perfect predictions about the world without any agreement on the interpretation of their theory.

In recent years a large body of theory has been developed under the name of quantum probability. This sophisticated mathematical theory, which certainly has much to say about quantum physics itself, is largely inspired by a powerful analogy between the mathematics, in quantum theory, of so-called states and observables, and the mathematics of probability measures and random variables as studied in ordinary probability theory. One can consider quantum probability as a generalisation of ordinary probability obtained by dropping the requirement of commutativity of certain objects. This reflects the physical fact that measurement of a quantum system disturbs that system; its state changes as a result of the measurement and the sequence in which one performs certain measurements influences strongly the outcomes one will obtain. Quantum probability theory has however not traditionally been much interested in the 'ordinary' probability theory of the outcomes of measurements, but rather has studied the internal evolution of a quantum system, and exploiting mathematical analogy has brought into the quantum world generalisations of classical notions from stochastic processes such as Markov processes, diffusions, and stochastic differential equations. Consequently we will not have much to say on this kind of quantum probability theory, though we hope to give some introduction to some current developments in that field.

Contents

Chapter 1

Introduction

In the first section of this introductory chapter we will describe a famous quantum experiment, the Stern-Gerlach experiment measuring the spin of the electron. Then in the second section we will give some rather abstract definitions of quantum state and measurement for a so-called finite-dimensional quantum system. In the d-dimensional case, these definitions involve various $d \times d$ complex matrices. Our aim in the rest of the chapter is to bring these two apparently unconnected starting points together. In fact it turns out that the spin of the electron can be described by the simplest non-trivial case of the abstract theory, the case $d = 2$. An electon's spin is always 'up' or 'down', relative to whatever direction you measure it, just 2 possibilities. An electron is a so-called spin-half particle: for a spin-j particle, with $j = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots$ one takes $d = 2j + 1$.

Before bringing the experiment and the theory together in the case $d = 2$, it will be helpful in section 3 to learn some remarkable facts about the 2×2 complex matrices, showing how their structure is closely linked to the structure of orientations and rotations in real three-dimensional space. With these preparations done we can finally (in section 4) show how our abstract set-up can be applied to the Stern-Gerlach experiment.

The mathematical model for $d = 2$ applies to a huge number of completely different physical systems. As well as to the spin of electrons, it applies to the polarization of photons, and to the energy levels of a so-called two level system: an object at such low temperature that only its 'ground state' and a first 'excited state' are relevant. The orientations and rotations in three-dimensional space which are naturally associated with the *same* mathematical model for all these systems, correspond to 'real' orientations and rotations for the spin of a spin-half particle like an electron. However for other systems there may be no real directions involved at all (the two level system) or the connection with real directions might be different and incom-

plete (polarization, where 'horizontal' and 'vertical' play the role of 'up' and 'down'). In quantum computation the word *qubit* is used to represent an arbitrary two-dimensional quantum system in its role a carrier of information. The phrases 'spin half' and 'two level system' are also often used to denote an arbitrary two-dimensional quantum syste.

In subsequent chapters we will develop the theory more systematically, but many times coming back to the two-dimensional case for illustration. Chapter 2 gives a systematic study of the finite-dimensional case, while chapter 3 extends to the infinite-dimensional case.

1.1 The Stern-Gerlach experiment

It seems only possible to give circular definitions of 'what is a quantum system', and what is a measurement of a quantum system. So let us not try but instead start by discussing in as simple terms as possible an exemplary special case: measuring the spin of electrons in a Stern-Gerlach device. This simple example shows that it is difficult if not impossible to really visualise what goes on when a quantum system is measured. Fortunately, despite the impossibility of understanding quantum phenomena in classical terms, we will see that the mathematical model for these phenomena—at least, the most simple kinds of quantum phenomena—is very simple indeed.

In 1922 Stern and Gerlach carried out an experiment to determine the size of the magnetic moment of the electron. The electron was conceived of as spinning around an axis and would therefore behave 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 magnet 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 size of the component of spin of each atom (or electron) in the direction of the gradient of the magnetic field. The spin in this direction would not be altered at all 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 only take on two different values.

Changing the orientation of the magnet did not alter the fact that the beam is split into two components, equally far apart, and of equal size.

There is no way that a cloud of points in $\mathbb{R}^3$ (the magnetic moments of a population of electrons) can project onto just two equally distant points when one projects them onto a straight line in any direction whatsover. A first conclusion seems to be that the passage through the magnet has changed the spin of each electron.

Versions of this experiment have later been carried out on other particles. Depending on the particle under study a possibly different finite number of values of spin are observed. One says that an electron has spin half; more generally, particles of spin $k/2$ where k is some integer are split into $k+1$ equally spaced beams by a Stern-Gerlach device.

Returning to the electrons carried by silver atoms, let us suppose now the top beam emerging from the first, vertically oriented Stern-Gerlach device, is sent again through another magnet in the same orientation as the first. Now all the silver atoms emerge in the top beam. This suggests that the direction of spin of the silver atoms has been changed by passage through the first magnet but is no longer altered by the second. More generally, if the second magnet makes an angle α with the first, then the proportion of the top beam from the first magnet which comes out top of the second (i.e., with spin in the direction α) is $\cos^2(\alpha/2)$. Repeating the process many times, using different angles, the picture emerges that after any splitting the atoms in one of the emerging beams have no memory of their earlier past and have a state determined in a stochastic fashion just by the last selection procedure to which they were subjected.

As long as we only split beams non-physicists will not find this experiment very though-provoking, since a layperson is perhaps not surprised that passage through the magnet alters spin into one of two different values. It appears that we should think of this passage through the magnet as changing the spin state of the atoms according to some simple Markovian rules. However if we add the possibility of merging beams new and extraordinary things happen.

The following discussion for silver atoms and the Stern-Gerlach device is only a thought experiment, but for other particles such an experiment has been done. If two emerging beams from a Stern-Gerlach device are made to recombine by passing them through an exactly reversed apparatus the result would again be a single beam, indistinguishable from the original. In itself this may not appear surprising but a further thought experiment shows that there is something wrong with our idea of a memoryless change of state during each passage.

Suppose a second magnet is placed horizontally across the path of the atoms in the upper beam emerging from a first, vertical, magnet. The beam emerging from the second magnet splits into two halves, spin left and spin

right say. What happens if these two beams are recombined into one? And then passed through a vertically oriented magnet? The electrons seem to remember their past and all emerge in the top beam of the last (fourth) magnet. If the first magnet is removed (the one which inititally split the beam into spin up and spin down particles, of which only the spin up particles were further treated) then the result would be two outgoing beams.

The picture that the spin of the electron is altered while passing through the magnetic field in a Markovian (memoryless) way now seems untenable. Is there some complicated mechanism whereby the past history of the atom is remembered and sometimes taken account of? This doesn't turn out to be a fruitful avenue of thought. There is a radical alternative, impossible to comprehend in classical terms, but which does give the right predictions. As long as no interaction with the outside world has taken place each silver atom passes out of the magnet as a wave with components in both beams. When the beam is stopped by a screen interaction takes place with the many particles in the screen, irreversible and macroscopic changes take place, and during the course of this process the wave with two components in the two beams has to make up its mind whether to be a particle in the lower or the upper beam. Similarly when we select one of the two beams for further experimentation, each atom has to decide whether to have taken the one path or the other. But without the selection, both paths were taken.

A very similar story can be told about polarization of photons. Throughout history whether light should be considered as a wave or a particle phenomenon has oscillated as new kinds of data became available. The nineteenth century view and the view of classical physics is that light is a wave phenomenon and the experimental fact that sometimes it behaves like a stream of particles belongs to the twentieth. One can split a beam of photons by passing it through a crystal of calcite (iceland spar) into two parts of different polarization. The emergent beams can be selected. Also polarization filters behave as selectors of photons of a particular direction of polarization. Photodetectors can register the arrival of a single photon at a detector. We will return to a further description of polarization in later sections. The mathematical model of states and measurements will appear the same, except that the correspondence between directions in space and these mathematical objects is slightly different: rotating a Stern-Gerlach device through 180° reverses its effect, while rotating a polarization filter 180° leaves its effect unchanged (a 90° rotation reverses the effect).

The above story is undoubtedly confusing. However let us draw from it the following conclusions. Particles produced in a laboratory, or arriving at a measurement apparatus from some external source, have a state. Measurement involves irreversible interaction of those particles with macroscopic

apparatus and produces outcomes which are random. From the state of the particles we will be able to compute the probability distribution of the outcome of any particular measurement which we may carry out. One should think of the state as a summary characteristic of how the particles were prepared, on the basis of which predictions of the future can be made. Whether a state can be ascribed to one particular particle or only to a potential stream of similarly prepared particles is a topic of debate in the foundational literature. It is connected to the similarly fundamental question of what one means by probability. We will say that a particular particle has, or is in, a particular state and use this to compute probabilities of particular outcomes when a measurement of just that particle is made. Unless the probability of some event happens to be zero or one, one can empirically interpret the prediction as being a statement about relative frequencies in a large sequence of identical repetitions of the experiment. It is in this sense that quantum mechanics is empirically tested in the laboratory.

Some measurements are selectors—we can separate particles according to the outcome of the measurement. In that case the state of the particles after measurement has changed and depends on the selection.

Measurement and selection involve irreversible interaction of the quantum system with the outside world. As electrons pass through a succession of Stern-Gerlach devices no measurement takes place—only when particles are absorbed or selected by a screen blocking all or some emergent beams can one say that measurement or selection has taken place. Which measurement has been taken depends on the whole configuration of devices prior to this interaction.

We will mainly be concerned with the mathematical model of measurement and the resulting statistical problems. That means that we will introduce mathematical objects to represent (a) the state of a quantum system, and (b) the measurement device. Some measurements can be also used as selectors in order to prepare a particle in a particular state.

Also of interest is how a quantum system evolves in isolation from the outside world—what happens between measurements, one could say. This evolution is deterministic and follows a so-called Schrödinger equation. We will see that it provides some interesting statistical models, as well as ways of building new measurement procedures from old. The reconciliation of these two different ways of evolution of a quantum system (random selection by measurement versus deterministic Schrödinger evolution) is called the measurement problem, a major topic in the foundations of quantum mechanics. Differing schools of thought exist, strongly coloured by the question whether one should see physics as merely giving recipes for succesful predictions of observable phenomena, in which case only an issue of self-consistency arises,

or whether one sees physics as giving a picture of an objective underlying reality, in which case that underlying reality is a very strange one indeed if quantum mechanics is to be fully believed.

1.2 The basic model

We now give a compact mathematical description of measurement of finite-dimensional quantum systems, as used to describe for example spin and polarization (as long as the spin or polarization can be considered in isolation from other aspects of the physical system concerned). On the other hand, to study position and momentum would require infinite-dimensional spaces and the mathematics becomes more delicate. However the main definitions are identical: the word matrix should then be replaced by the word operator, and the space $\mathbb{C}^d$ by an arbitrary separable complex Hilbert space $\mathcal{H}$. In Chapter 3 we will provide the technical details of this generalization.

1.2.1 States

Definition 1 (State, density matrix, density operator). The *state* of a d-dimensional quantum system is specified by a $d \times d$ complex matrix ρ, called the *density matrix* or density operator of the system. A density matrix is *self-adjoint*, *nonnegative*, and has *trace* equal to 1.

The technical terms in this definition (self-adjoint, nonnegative, trace) will be explained in a moment. To illustrate the choice of dimension, we note that the state of spin of one electron (a so-called spin half particle), can be described with $d = 2$. More generally a spin $k/2$ particle needs a description with $d = k + 1$. Photons are actually spin 1 particles but the polarization of one photon can be modelled with $d = 2$. As we will later see, the state of a system of n spin half particles will be described with $d = 2^n$.

Definition 2 (Self-adjoint matrix, operator). Using as usual a bar to denote complex conjugation and the symbol $\top$ to denote transpose of matrix, the *adjoint* of a matrix is simply the transpose and element-wise complex conjugate of the matrix. So saying that ρ is self-adjoint is saying that $\rho_{ij} = \overline{\rho}_{ji}$ for all $i, j = 1, \ldots, d$. Using a star to denote taking the adjoint, i.e., transposition and complex-conjugation, we can write this as $\rho = \rho^* = \overline{\rho}^\top$.

When we extend to the infinite-dimensional case, matrices will be replaced by operators. Some authors use the word Hermitian instead of self-adjoint, and a dagger instead of a star. But be warned, for other authors Hermitian and self-adjoint mean different things.

Definition 3 (Nonnegative matrix, operator). One may check that if a matrix ρ is self-adjoint then for any complex column vector ψ, the quadratic form $\psi^*\rho\psi$ is a real number. When we say that ρ is *nonnegative* we mean that $\psi^*\rho\psi \geq 0$ for all vectors ψ.

Definition 4 (Bra's and ket's). We will often use the Dirac *bra-ket* notation whereby we write $|\psi\rangle$ for the complex column vector ψ and $\langle\psi|$ for the complex row vector ψ^*. Thus $\langle\psi \mid \psi\rangle$ is the real scalar quantity $\psi^*\psi$, the inner product of ψ with itself or its squared length, while $|\psi\rangle\langle\psi|$ is the self-adjoint, rank one, matrix $\psi\psi^*$. Instead of $\psi^*\rho\psi$ one would write $\langle\psi \mid \rho \mid \psi\rangle$ and so on.

Definition 5 (Trace). The trace of a matrix is the sum of its diagonal elements.

The diagonal elements of a self-adjoint matrix are real numbers.

The key fact we need about self-adjoint matrices is that they have real eigenvalues, and eigenvectors which can be taken to form an orthonormal basis of $\mathbb{C}^d$. Thus any density matrix can be written as $\rho = \sum_{i=1}^d p_i\psi_i\psi_i^*$, or if you prefer $\rho = \sum_{i=1}^d p_i |\psi_i\rangle\langle\psi_i|$ where the p_i are eigenvalues and the ψ_i are corresponding orthornormal eigenvectors. The latter notation will often be abbreviated for instance to $\sum_{i=1}^d p_i |i\rangle\langle i|$. The trace of a self-adjoint matrix is equal to the sum of its eigenvalues and a nonnegative self-adjoint matrix has nonnegative eigenvalues. Hence the eigenvalues p_i are nonnegative and add up to one—they form a probability distribution over the d eigenvectors ψ_i, $i = 1, \ldots, d$. In fact each of the d matrices $\psi_i\psi_i^*$ is itself a density matrix. One says that the density matrix ρ is the mixture of the *pure states* $\psi_i\psi_i^*$, sometimes referred to just by the vectors ψ_i. As we will later see, if one prepares with probability p_i a quantum system in the pure state ψ_i, the result is indistinguishable from preparing the system in the state $\rho = \sum_{i=1}^d p_i\psi_i\psi_i^*$.

1.2.2 Measurements

The result of measurement of a quantum system could be an integer, a real number, a vector of numbers. In the Stern-Gerlach experiment choosing a particular orientation of the magnet and letting one particle pass through it, then observing whether it takes the top or the bottom route, is a measurement whose outcome only takes two values which we might arbitrarily label $+1$, -1. But we emphasize that each different orientation of the magnet is a different measurement, a different experiment which one could do, whether or not we use the same or different labels for the two outcomes of each possible experiment. In that example experiments are possible taking more different

values. For instance, repeatedly splitting a beam of spin half particles by n Stern-Gerlach devices into 2^n possible output beams gives a measurement, on one particle, with 2^n different outcomes. Merging some of the output beams would give measurements with a smaller number of different outcomes. Finally, suppose one chooses an angle uniformly at random between 0 and 2π and lets a particle pass through one Stern-Gerlach device oriented at that angle, and notes the angle and ± 1 according to the route taken by the particle. This is now a measurement taking values in $[0, 2\pi) \times \{-1, +1\}$. Obviously it can be described in terms of ordinary randomisation and one of the simple measurements first considered, but still it constitutes in itself a measurement with values in the space mentioned. As we shall see later there are yet more possibilities.

It turns out (and we will also explain later why) that the broadest possible class of measurements can be represented mathematically with the class of *operator-valued probability measures*, or OProM's as we shall abbreviate them. Sometimes one also uses the name 'generalised measurement', and sometimes the abbreviation 'povm' is used (positive operator valued measure). Within that huge class of all conceivable measurements there is a special subclass of measurements, called simple or von Neumann measurements, and represented by *projector-valued probability measures* or ProProM's. In the Stern-Gerlach example, a single Stern-Gerlach device corresponds to a simple measurement, but any combination of any number of devices—and yet more complicated experiments too—will be modelled by generalized measurements. Here are the definitions:

Definition 6 (Measurement). Let $(\mathcal{X}, \mathcal{B})$ be a measurable space. A *(generalised) measurement* or OProM on a d-dimensional quantum system, taking values in the space $\mathcal{X}$, can be represented by a collection M of $d \times d$ self-adjoint complex matrices $M(B) : B \in \mathcal{B}$ forming an *operator-valued probability measure*, i.e., satisfying the following three conditions:

1. each $M(B)$ is non-negative.
2. for any countable collection of pairwise disjoint B_i with $B = \cup_i B_i$ we have $M(B) = \sum M(B_i)$.
3. $M(\mathcal{X}) = \mathbf{1}$, the identity matrix.

Note that these three conditions are the usual axioms of a probability measure except that the values are self-adjoint operators, not real numbers. If the sample space $\mathcal{X}$ is countable then just as in ordinary probability theory it is more convenient to describe a measurement with values in $\mathcal{X}$ just with the values $M(\{x\})$ which we would notationally abbreviate to $M(x)$. Then

each $M(x)$ is self-adjoint and nonnegative, and they add up to the identity $\mathbf{1}$.

Definition 7 (Simple Measurement). A *simple* or von Neumann measurement or ProProM is a measurement for which each $M(B)$ is a projector, i.e., a matrix which projects onto a subspace of $\mathbb{C}^d$. Equivalently, alongside being self-adjoint and nonnegative, each $M(B)$ is idempotent, $M(B)^2 = M(B)$.

Much more can be said (and later will be said) concerning properties of OProM's and ProProM's, or (generalised) measurements and simple measurements respectively, but let us complete the outline by now giving the so-called *trace rule* assigning to the combination of a quantum state and a measurement, the probability distribution of the outcome of the measurement. The outcome is random and takes values in $\mathcal{X}$. Let us denote it by a random variable X. Then we have

$$\Pr\{X \in B\} = \mathrm{trace}(\rho M(B)) \qquad \text{for all } B \in \mathcal{B}. \tag{1.1}$$

The reader should verify at this point that this does indeed specify a probability measure on $(\mathcal{X}, \mathcal{B})$! Note that if $M(A) = \mathbf{1}$ for some $A \subseteq \mathcal{X}$, then the outcome X lies with probability 1 in the set A.

We can now define *quantum statistics*, or quantum statistical inference. Suppose the state ρ of a quantum system depends on an unknown parameter θ (typically a real vector taking values in a parameter space Θ; so we write $\rho = \rho(\theta)$ and in (1.1) the distribution of X, the outcome of measurement of M, depends on the unknown θ. Can we identify interesting quantum statistical problems for which natural physical properties of the model lead to special or attractive statistical features of the inference problem? What can be said about which measurement should be taken in order to get most information about θ? How can one implement a desired measurement M? Sometimes the measurement to be taken in practice has some unknown features too, e.g., noise levels or imperfect detection rates, and then not only ρ but also (or instead) M depends on unknown parameters.

We can deduce a lot more about the nature of a *simple* measurement. The sum of two projectors is only itself a projector if the corresponding two subspaces are orthogonal. Therefore for disjoint B_1, B_2, it must be that $M(B_1)$ and $M(B_2)$ project onto orthogonal subspaces and $M(B_1 \cup B_2) = M(B_1) + M(B_2)$ projects onto the sum space generated by these two subspaces. In the d-dimensional case considered at the moment, one can see that a simple measurement takes on at most d different values, i.e., for some $x_1, \ldots, x_k$ with $k \leq d$, $M(\{x_1, \ldots, x_k\}) = \mathbf{1}$ and hence the measurement is

characterized just by the projectors $M(x_i) = M(\{x_i\})$ or equivalently by the k orthogonal subspaces, together spanning $\mathbb{C}^d$, onto which they project.

Simple measurements can be used to prepare a quantum system in a certain state, since there is a rather simple rule which tells us the state of the system *after* measurement has yielded the value x: namely the system is now in the state $M(x)\rho M(x)/\operatorname{trace}\rho M(x)$. This rule is called the *projection postulate*, and often associated with the names of von Neumann and Lüders. For generalised measurements the situation is much more complicated, and we will postpone discussion of that till chapter 2.

1.3 Spin half

A two-dimensional quantum system is the simplest we can study. It has important applications as well as already very rich structure. Furthermore, there is an elegant geometric picture of states and simple measurements in terms of directions in three dimensional real space. This will allow us to visualise a correspondence between the abstract notions of state and measurement on the one hand, and real physical systems on the other. The key applications are to the spin of electrons (or other spin half particles), and the polarization of photons.

1.3.1 The geometry of spin half

First we take some time to study some special features of the 2×2 self-adjoint matrices. The properties we find will greatly simplify calculations. Let $\mathbf{1}$ denote the identity matrix and define the *Pauli spin matrices* as follows:

Definition 8 (Pauli spin matrices).

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.2}$$

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

$$\begin{aligned} \sigma_x\sigma_y = -\sigma_y\sigma_x &= i\sigma_z, \\ \sigma_y\sigma_z = -\sigma_z\sigma_y &= i\sigma_x, \\ \sigma_z\sigma_x = -\sigma_x\sigma_z &= i\sigma_y, \\ \sigma_x^2 = \sigma_y^2 = \sigma_z^2 &= \mathbf{1}. \end{aligned} \tag{1.3}$$

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

$$X = \begin{pmatrix} u+z & x-iy \\ x+iy & u-z \end{pmatrix} \tag{1.4}$$

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

$$X = u\mathbf{1} + x\sigma_x + y\sigma_y + z\sigma_z \tag{1.5}$$

and the representation is unique. This shows that the space of all 2×2 self-adjoint (complex) matrices can actually be thought of as a four-dimensional real vector space.

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

$$\rho = \rho(\vec{a}) = \frac{1}{2}(\mathbf{1} + \vec{a} \cdot \vec{\sigma}) \tag{1.6}$$

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

$$\|\vec{a}\|^2 = a_x^2 + a_y^2 + a_z^2 \leq 1 \tag{1.7}$$

while $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and '$\cdot$' denotes the inner-product. Thus the space of density matrices of a two-dimensional quantum system can be represented by the closed unit ball in three dimensional Euclidean space. The sphere, or surface of the unit ball, then corresponds to density matrices $\frac{1}{2}(\mathbf{1} + \vec{a} \cdot \vec{\sigma})$ with $\|\vec{a}\|^2 = 1$ which are singular since their determinant is zero. Such a density matrix has therefore eigenvalues 0 and 1 and hence is a projector matrix, projecting onto a one-dimensional subspace of $\mathbb{C}^2$. Letting $\vec{u}$ denote a unit vector in $\mathbb{R}^3$, let us write $\Pi(\vec{u}) = \frac{1}{2}(\mathbf{1} + \vec{u} \cdot \vec{\sigma})$ for this matrix. Check using (1.3) that $\Pi(\vec{u})$ is idempotent! And check that $\Pi(\vec{u})$ and $\Pi(-\vec{u})$ commute (in fact, their product is the zero matrix) and add to the identity matrix! Thus the projectors $\Pi(\vec{u})$ and $\Pi(-\vec{u})$ project onto two orthogonal one-dimensional subspaces of $\mathbb{C}^2$. We will determine these spaces exactly in a moment.

In fact, the only other projector matrices are the zero matrix $\mathbf{0}$ and the identity matrix $\mathbf{1}$, which are of rank 0 and rank 2 respectively.

It follows that for an arbitrary density matrix $\rho = \rho(\vec{a})$, defining the unit vector $\vec{u} = \vec{a}/\|\vec{a}\|$, we have

$$\begin{aligned} \rho(\vec{a}) = \tfrac{1}{2}(\mathbf{1} + \vec{a} \cdot \vec{\sigma}) &= \tfrac{1}{2}(1 + \|\vec{a}\|)\rho(\vec{a}/\|\vec{a}\|) + \tfrac{1}{2}(1 - \|\vec{a}\|)\rho(-\vec{a}/\|\vec{a}\|) \\ &= \tfrac{1}{2}(1 + \|\vec{a}\|)\Pi(\vec{u}) + \tfrac{1}{2}(1 - \|\vec{a}\|)\Pi(-\vec{u}). \end{aligned} \tag{1.8}$$

It has eigenvalues $\frac{1}{2}(1+\|\vec{a}\|)$ and $\frac{1}{2}(1-\|\vec{a}\|)$, and its eigenvectors, column vectors in $\mathbb{C}^2$, generate the spaces onto which $\Pi(\vec{u})$ and $\Pi(-\vec{u})$ project.

So what are these spaces exactly? The vector $\vec{u}$ is a point on the unit sphere in $\mathbb{R}^3$. Let θ and ϕ denote its polar coordinates, where $\theta \in [0,\pi)$ is the latitude measured from the North pole (z-axis) and $\phi \in [0,2\pi)$ is the longitude, measured from the x-axis. Thus $\vec{u} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. Define the column vector $|\psi\rangle = |\psi(\theta,\phi)\rangle$ in $\mathbb{C}^2$ by

$$|\psi(\theta,\phi)\rangle = \begin{pmatrix} e^{-i\phi/2}\cos(\theta/2) \\ e^{i\phi/2}\sin(\theta/2) \end{pmatrix}. \tag{1.9}$$

Note that $\langle\psi \mid \psi\rangle = 1$ while

$$\begin{aligned} |\psi\rangle\langle\psi| &= \begin{pmatrix} e^{-i\phi/2}\cos(\theta/2) \\ e^{i\phi/2}\sin(\theta/2) \end{pmatrix} \begin{pmatrix} e^{i\phi/2}\cos(\theta/2) & e^{-i\phi/2}\sin(\theta/2) \end{pmatrix} \\ &= \begin{pmatrix} \cos^2(\theta/2) & e^{-i\phi}\cos(\theta/2)\sin(\theta/2) \\ e^{i\phi}\cos(\theta/2)\sin(\theta/2) & \sin^2(\theta/2) \end{pmatrix} \\ &= \tfrac{1}{2}\begin{pmatrix} 1+\cos(\theta) & (\cos\phi - i\sin\phi)\sin\theta \\ (\cos\phi - i\sin\phi)\sin\theta & 1-\cos\theta \end{pmatrix} \\ &= \tfrac{1}{2}(\mathbf{1} + \vec{u}\cdot\vec{\sigma}) = \Pi(\vec{u}). \end{aligned} \tag{1.10}$$

Note that *any* complex vector $|\xi\rangle$ of length 1 can be written as $e^{i\alpha}\psi(\theta,\phi)$ for some $\alpha \in [0,2\pi)$ and polar coordinates θ,ϕ. Note that $|\xi\rangle\langle\xi| = |\psi\rangle\langle\psi| = \Pi(\vec{u})$. And that $|\psi(\theta,\phi)\rangle$ and $|\psi(\pi-\theta,\phi+\pi)\rangle$ are orthogonal. The corresponding points on the unit sphere are opposite to one another. Combining these facts we obtain:

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

Let $\vec{u}$ and $\vec{v}$ be two unit vectors in $\mathbb{R}^3$ and write $|\vec{u}\rangle$ and $|\vec{v}\rangle$ for the corresponding unit vectors in $\mathbb{C}^2$; so $|\vec{u}\rangle$ is an abbreviation for $|\psi(\theta,\phi)\rangle$ where θ,ϕ are the polar coordinates of $\vec{u}$. Since $\Pi(\vec{u}) = |\vec{u}\rangle\langle\vec{u}|$ we see that $\operatorname{trace}\Pi(\vec{u})\Pi(\vec{v}) = \operatorname{trace}|\vec{u}\rangle\langle\vec{u}|\,|\vec{v}\rangle\langle\vec{v}| = \langle\vec{v}\mid\vec{u}\rangle\langle\vec{u}\mid\vec{v}\rangle = |\langle\vec{u}\mid\vec{v}\rangle|^2$. On the other hand, using the properties (1.3) of the Pauli matrices, one can compute (and the reader should do this for him or herself!) $\operatorname{trace}\Pi(\vec{u})\Pi(\vec{v}) = \frac{1}{2}(1+\vec{u}\cdot\vec{v})$. Now $\vec{u}\cdot\vec{v}$ is the cosine of the angle between the vectors $\vec{u}$ and $\vec{v}$, hence $\frac{1}{2}(1+\vec{u}\cdot\vec{v})$ is the squared cosine of half the angle between $\vec{u}$ and $\vec{v}$.

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

With these geometric and algebraic facts on record, we now return to measurement of a spin half system.

1.3.2 Measurement for spin half

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

What have we learnt from the previous subsection about the classes of simple and generalised measurements, ProProM's and OProM's? For the simple measurements, the description is indeed very simple.

Consider a simple measurement M taking values in a set $\mathcal{X}$ consisting of just two elements, let's call these elements ± 1. We need to associate a projector $M(B)$ with B equal to each of $\emptyset$, $\{+1\}$, $\{-1\}$, and $\mathcal{X} = \{+1, -1\}$, such that the sums of the projectors for disjoint sets B is the projector for their union, and the projector for the whole set $\mathcal{X}$ is the identity matrix. This forces $M(\emptyset) = \mathbf{0}$, $M(\mathcal{X}) = \mathbf{1}$, projecting onto the zero-dimensional subspace $\{0\}$ and the two-dimensional subspace $\mathbb{C}^2$ respectively. One-dimensional subspaces of $\mathbb{C}^2$ are each generated by a vector of the form $|\vec{u}\rangle$ for some $\vec{u}$ on the unit sphere, and the associated projectors are $\Pi(\vec{u})$. Recall that opposite points $\pm\vec{u}$ on the unit sphere correspond to orthogonal vectors $|\pm\vec{u}\rangle$ in $\mathbb{C}^2$, and hence to orthogonal projectors $\Pi(\pm\vec{u})$. Thus a particular projector-valued probability measure for a measurement with values in $\mathcal{X}$ is given by $M(\{\pm 1\}) = \Pi(\pm\vec{u}) = \frac{1}{2}(\mathbf{1} \pm \vec{u} \cdot \vec{\sigma})$ for some $\vec{u}$.

Some thought will show that even if $\mathcal{X}$ has more elements than just two, really every ProProM is equivalent to the one we have just described, or otherwise to a trivial ProProM only taking a single value. Let us continue by using the trace rule (1.1) to compute the probabilities of the two outcomes ± 1 when the measurement $M(\{\pm 1\}) = \Pi(\pm\vec{u})$ is carried out on a system in the state $\rho(\vec{a}) = \frac{1}{2}(\mathbf{1} + \vec{a} \cdot \vec{\sigma})$. Using the properties (1.3) of the Pauli matrices, the reader should verify that these probabilities are

$$\operatorname{trace} \rho(\vec{a})\Pi(\pm\vec{u}) = \frac{1}{2}(1 \pm \vec{a} \cdot \vec{u}). \tag{1.11}$$

What is the state of the system after the measurement has yielded the value ± 1? We must compute $M(\{\pm 1\})\rho M(\{\pm 1\}) = \Pi(\pm\vec{u})\rho(\vec{a})\Pi(\pm\vec{u})$ and

divide by its trace, which is just the probability (1.11). The first step yields

$$\Pi(\vec{u})\rho(\vec{a})\Pi(\vec{u}) = |\vec{u}\rangle\langle\vec{u}|\,\rho\,|\vec{u}\rangle\langle\vec{u}| = \langle\vec{u}\mid\rho\mid\vec{u}\rangle\,|\vec{u}\rangle\langle\vec{u}| = \mathrm{trace}(\rho\,|\vec{u}\rangle\langle\vec{u}|)\,|\vec{u}\rangle\langle\vec{u}| \tag{1.12}$$

which is just the density matrix $\rho(\vec{u}) = \Pi(\vec{u})$ times the corresponding probability (1.11).

Thus after measurement has yielded the value ± 1, the system is in the *pure state* specified by the vector $|\pm\vec{u}\rangle$.

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

A Stern-Gerlach magnet oriented in the direction $\vec{u}$ implements the simple measurement $M(\{\pm 1\}) = \Pi(\pm\vec{u})$. Since for $\vec{a} = \vec{0}$ the probabilities (1.11) both equal $\frac{1}{2}$, one will find electrons with spin in the directions $\pm\vec{u}$ with equal probabilities. Electrons in the emerging '+' beam are in the pure state $\rho(\vec{u})$. Sending them through a Stern-Gerlach device with orientation $\vec{v}$ splits them again, now with probabilities $\frac{1}{2}(1 \pm \vec{u}\cdot\vec{v})$ into two beams of electrons in the states $\rho(\pm\vec{v})$ and so on. Recall that $\frac{1}{2}(1 \pm \vec{u}\cdot\vec{v})$ is the squared cosine of half the angle between the directions $\vec{u}$ and $\vec{v}$.

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

If we are allowed to measure a large number of electrons each in the same mixed state $\rho(\vec{a})$, we see that a large number of Stern-Gerlach measurements in three linearly independent directions will determine $\vec{a}$.

At this stage we will not say much about generalized measurements. The class of all generalized measurements is very large, and to explain how the more interesting ones might be realised, we will need to go further into the theory. However one can build some generalised measurements by carrying out a succession of simple ones, or choosing simple ones at random. The problem section at the end of this chapter give some examples.

1.3.3 Polarization.

Polarization is the movement in an ellipse in the plane prependicular to the direction of propagation of an electromagnetic wave, e.g., light. Map ellipses to sphere: circles clockwise, anticlockwise at North, South pole; ellipse (clockwise on Northern hemisphere, anticlockwise on Southern) collapses to line in each possible direction on equator. Rotation of ellipse by 180° restores it to original. Polarization filters and simple measurements. For other possibilities see Peres. Ammonia maser—another example of a two state system.

1.4 Problems and further results

Problem 1. Show that the trace rule (1.1) does indeed define a probability distribution on $(\mathcal{X}, \mathcal{B})$.

Problem 2. Show that a simple measurement on a d-dimensional quantum system takes on at most d different values.

Problem 3. Show that the matrices $M(B)$ of a simple measurement are idempotent and commute, $M(B_1)M(B_2) = M(B_2)M(B_1) = M(B_1 \cap B_2)$.

Problem 4. Verify the properties (1.3) of the Pauli spin matrices.

Problem 5. Show that $(\vec{a} \cdot \vec{\sigma})(\vec{b} \cdot \vec{\sigma}) = (\vec{a} \cdot \vec{b})\mathbf{1} + i(\vec{a} \times \vec{b}) \cdot \sigma$

Problem 6. Verify that $\operatorname{trace} \rho(\vec{a})\Pi(\vec{u}) = \frac{1}{2}(1 + \vec{a} \cdot \vec{u})$.

Problem 7. Suppose that $\vec{u}_1$, $\vec{u}_2$, $\vec{u}_3$ are the results of a rotation in $\mathbb{R}^3$ acting on the unit vectors in the x, y and z directions respectively. Show that the three matrices $\sigma_i = \vec{u}_i \cdot \vec{\sigma}$ satisfy the same properties (1.3) as the Pauli matrices themselves.

Problem 8. Define the matrix $U = e^{-i\eta \vec{u} \cdot \vec{\sigma}/2}$ by formal use of the power series expansion of the exponential function. Show that $U = \cos(\eta/2)\mathbf{1} - i \sin(\eta/2)\vec{u} \cdot \vec{\sigma}$. Show that $U\rho(\vec{a})U^*$ transforms the state $\rho(\vec{a})$ into the state $\rho(\vec{b})$ where $\vec{b}$ results from $\vec{a}$ by rotation about the direction $\vec{u}$ through an angle θ. Hint: check the result first for the case when $\vec{u} = (0, 0, 1)$ and when $\rho(\vec{a})$ is the pure state $|\psi(\theta, \phi)\rangle \langle\psi(\theta, \phi)|$. Next use the result of the previous problem to convert the general case into this special case.

Problem 9. Suppose that electrons are to be sent first through a Stern-Gerlach device with orientation $\vec{u}$, and then that each emergent beam is sent through devices with orientiation $\vec{v}$, so that four possible output beams

result. Use this device to make a measurement on an electron with four possible outcomes $\pm\pm$. Show that this is a generalised measurement with $M(\pm_1\pm_2) = \Pi(\pm_1\vec{u})\Pi(\pm_2\vec{v})\Pi(\pm_1\vec{u})$.

Problem 10. Suppose that a Stern-Gerlach device is oriented in a random direction $\vec{u}$, uniformly distributed on the sphere, and then used to measure spin resulting in $\pm$. Thus we have a measurement taking values $\theta, \phi, \pm$ where θ, ϕ are the polar coordinates of $\vec{u}$. Show that this measurement corresponds to the OProM $M(\mathrm{d}\theta, \mathrm{d}\phi, \pm) = \Pi(\pm\vec{u}(\theta,\phi))\sin\theta \mathrm{d}\theta\mathrm{d}\phi/4\pi$.

Problem 11. Show that $M(\mathrm{d}\theta, \mathrm{d}\phi) = \Pi(\vec{u}(\theta,\phi))\sin\theta\mathrm{d}\theta\mathrm{d}\phi/4\pi$ defines another OProM.

Problem 12. Compute the Fisher information matrix for the unknown parameter $\vec{a}$ based on either of the measurements of the last two problems.

1.5 Bibliographic comments

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

The classic books by Helstrom (1976) and Holevo (1982) are still the only books on quantum statistics and they are very difficult indeed to read for a beginner. A more recent resource is the survey paper by Malley and Hornstein (1993). However the latter authors, among many distinguished writers both from physics and from mathematics, take the stance that the randomness occurring in quantum physics cannot be caught in a standard Kolmogorovian framework. We argue elsewhere (Gill, 1998), in a critique of an otherwise excellent introduction to the related field of quantum probability (Kümmerer and Maassen, 1998), that depending on what you mean by such a statement, this is at best misleading, and at worst simply untrue.

See Penrose (1994) for an eloquent discussion of why there is something paradoxical in the peaceful coexistence of the two principles of the deterministic, unitary evolution of an isolated quantum system, and the random jump which occurs when it is measured. See Percival (1998) for interesting stochastic modifications to Schrödinger's equation which might offer some reconciliation. Also highly recommended: Sheldon Goldstein, 'Quantum mechanics without observers', Physics Today, March, April 1998; letters to the editor, Physics Today, February 1999.

Chapter 2

Observables and wave functions

In Chapter 1 we stated that the state of a quantum system was specified by a density matrix ρ, while a measurement on that system is specified by an operator-valued probability measure. We noted the existence of a special kind of state, namely the *pure* states for which ρ is a rank one projector; and a special kind of measurement, called a simple measurement, specified by a *projector*-valued probability measure. In this chapter we will start by considering these special states and special measurements in more detail. For both these objects a more compact description is available. A density matrix ρ of rank 1 is of the form $\rho = |\psi\rangle\langle\psi|$ for some vector $|\psi\rangle$ of length 1, i.e., such that $\langle\psi \mid \psi\rangle = 1$. The *state vector* or *wave-function* ψ is unique up to multiplication by a complex number of absolute value 1. As for the simple measurements, recall that a simple measurement on a finite-dimensional state, of dimension d, is characterized by $k \leq d$ distinct values $x_1, \ldots, x_k$ and projectors $M(x_1), \ldots, M(x_k)$, the latter projecting onto orthogonal subspaces together spanning $\mathbb{C}^d$. Without loss of generality we may assume that none of these projectors is the trivial projector $\mathbf{0}$; otherwise, simply delete it and the corresponding x_i from our list. Whatever the state of the system, the probability will be 0 to obtain this value. Write Π_A for the projector onto the subspace A and suppose that $M(x_i) = \Pi_{A_i}$. Suppose the sample space $\mathcal{X}$ is the real line or a subset thereof, so that the x_i are real numbers. Then we note that $X = \sum_i x_i \Pi_{A_i}$ is a self-adjoint operator. Its eigenvalues are $x_1, \ldots, x_k$ and the corresponding eigenspaces are $A_1, \ldots, A_k$. On the other hand, any self-adjoint operator is of this form. Thus there is a one-to-one correspondence between real-valued simple measurements, and self-adjoint operators. When thinking of a self-adjoint operator X as a description of a simple measurement, we call it an *observable*. The associated projector-valued probability measure is then also called its *spectral decomposition*.

2.1 Observables and functions of observables

2.1.1 Obervables

Let X be an observable on a d-dimensional quantum system. Let us denote by $[X = x]$ the eigenspace of X corresponding to eigenvalue x; that is

$$[X = x] = \{\psi : X\psi = x\psi\}. \tag{2.1}$$

Thus if x is not actually an eigenvalue of X, $[X = x]$ is the trivial subspace $\{0\}$. For a Borel set B on the real line, write $[X \in B]$ for the sumspace generated by all $[X = x]$ for $x \in B$. Write $\Pi_{[X \in B]}$ for the projector onto this space. Note that $(\Pi_{[X \in B]} : B \in \mathcal{B})$ is a projector-valued probability measure with values in $(\mathbb{R}, \mathcal{B})$.

We can write

$$X = \sum_x x \Pi_{[X=x]} \tag{2.2}$$

where we need only sum over x in the spectrum of X, the set of all its eigenvalues.

2.1.2 Decomposition of $\mathbb{C}^d$

A collection $A_1, \ldots, A_k$ of subspaces of $\mathbb{C}^d$ is called a decomposition of $\mathbb{C}^d$ if they are orthogonal and their sumspace is all of $\mathbb{C}^d$. The decomposition is called proper if none of the A_i are the trivial subspace $\{0\}$. There is thus a one-to-one correspondence between proper decompositions of $\mathbb{C}^d$ each element of which is labelled with a distinct real value, and observables.

2.1.3 Function of an observable

If f is a real function and X is an observable, then by $f(X)$ we denote the new observable

$$f(X) = \sum_x f(x) \Pi_{[X=x]}. \tag{2.3}$$

In other words, the eigenvectors remain the same but the eigenvalues are replaced by f of the original eigenvalues. If f is not one-to-one then some eigenspaces of X may be merged together into eigenspaces of $f(X)$. This definition agrees with the more obvious (algebraic) definition of such functions of X as $aX + b$, X^2, X^{-1} (if X is non-singular), and so on. All functions of one observable commute, i.e., $f(X)g(X) = g(X)f(X)$.

Taking for f the indicator function 1_B of a Borel set B we note that $1_B(X) = \Pi_{[X \in B]}$.

A projector Π_A is an observable, with spectrum just the two values 1 and 0, and with the subspaces A and its orthocomplement $A^\perp$ being the eigenspaces.

2.1.4 Compatibility

We say that two subspaces A, A' of $\mathbb{C}^d$ are compatible if $A = (A \cap A') \oplus (A \cap A'^\perp)$ and vice versa for A'. Equivalently, the projectors Π_A and $\Pi_{A'}$ commute.

We say that two observables are compatible if and only if as operators they commute. In this case each pair of subspaces $[X = x]$ and $[Y = y]$ is compatible and the subspaces $[X = x] \cap [Y = y]$, which we abbreviate to $[X = x, Y = y]$, form a decomposition of $\mathbb{C}^d$ as x and y vary throughout the spectra of X and Y. Equivalently, X and Y possess a common orthonormal basis of eigenvectors. If X and Y commute they are both functions of a third observable Z; this can for instance be defined by giving a distinct label z to each of the non-zero subspaces $[X = x] \cap [Y = y]$.

2.1.5 Function of several compatible observables

We can also define functions of two (or more) *commuting* observables by

$$f(X, Y) = \sum_x \sum_y f(x, y) \Pi_{[X=x, Y=y]}. \tag{2.4}$$

Again, $f(X, Y)$ commutes with X and Y. Taking f to be the sum or product gives an alternative definition of $X + Y$ and XY, restricted to the case that X and Y commute. Fortunately it agrees with the usual algebraic definition of the sum or product of two operators (in our case, matrices) for which of course no assumption of commutativity is needed.

Example 1 (spin half). In the spin half case ($d = 2$), two observables X, Y commute if and only if they can both be written in the form $a\mathbf{1} + b\vec{u} \cdot \vec{\sigma}$ with the *same* $\vec{u}$ but with possibly different a, b. Thus any collection of commuting observables is a collection of functions of $\vec{u} \cdot \vec{\sigma}$ for some $\vec{u}$.

2.2 Measuring an observable on a pure state

2.2.1 Pure states

A quantum system in a pure state is characterized by its state-vector $|\psi\rangle$, an element of $\mathbb{C}^d$ of unit length, $\langle\psi \mid \psi\rangle = 1$. The corresponding density matrix is $|\psi\rangle\langle\psi|$ which can equivalently be characterised as the matrix which projects onto the one-dimensional subspace $[\psi] = \{c\,|\psi\rangle : c \in \mathbb{C}\}$. Multiplying $|\psi\rangle$ by a complex number of modulus 1 does not change the density matrix. As we will see, the choice of *phase* needed when we step over to the state-vector description is arbitrary—it does not affect any predictions we make.

There are two basic ways in which the state of a quantum system may change. If the quantum system is left in isolation from the outside world then its state changes deterministically according to the Schrödinger equation. However, if an observable is measured the state of the system makes an instantaneous random jump to a new state, corresponding to the outcome of the measurement.

2.2.2 Unitary evolution

A pure quantum state evolves in isolation from the outside world according to the Schrödinger differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\psi(t) = iH\psi(t) \tag{2.5}$$

where H is a self-adjoint operator called the *Hamiltonian*. (Usually H is written as $\hbar H$ where $\hbar$ is Planck's constant but we will absorb this factor into the Hamiltonian.) For instance, when we model a spin half particle in a magnetic field the Hamiltonian will depend on the strength and orientation of the magnetic field. In the finite-dimensional case the equation can be solved explicitly as

$$\psi(t) = e^{iHt}\psi(0). \tag{2.6}$$

Note that the matrix $U = e^{iHt}$ is *unitary*, that is, it satisfies $UU^* = U^*U = \mathbf{1}$. Consequently transformation of a collection of vectors ψ by multiplying by U does not change the lengths or the innerproducts of the vectors. One says that the state $\psi(t)$ follows a unitary (or Hamiltonian, or Schrödinger) evolution. For later reference note that equivalently the density matrix $\rho = |\psi\rangle\bar{\psi}$ satisfies the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = iH\rho(t) - i\rho(t)H = i[H, \rho(t)] \tag{2.7}$$

where the *commutator* of two observables X, Y is defined by $[X, Y] = XY - YX$. This equation has as solution

$$\rho(t) = e^{iHt}\rho(0)e^{-iHt}. \tag{2.8}$$

Example 2 (spin half). Consider the spin half case and suppose $\psi = \psi(0) = |\vec{u}\rangle$ while suppose $H = a\mathbf{1} + b\vec{v}\cdot\vec{\sigma}$. This corresponds to a magnetic field in the direction $\vec{v}$ of strength b. From problem 8 we see that that $e^{iHt}|\vec{u}\rangle\langle\vec{u}|e^{-iHt} = |\vec{u}(t)\rangle\langle\vec{u}(t)|$ where $\vec{u}(t) \in \mathbb{R}^3$ is the result of rotating $\vec{u}$ about $\vec{v}$ through the angle $-2bt$. Thus $\psi(t) = e^{i\theta(t)}|\vec{u}(t)\rangle$, and since we may ignore the phase $e^{i\theta(t)}$ (which we do not try to compute in general) the evolution of the system is that the direction of its spin $\vec{u}$ precesses around the direction of the magnetic field $\vec{v}$ at constant rate $-2b$. One can now imagine statistical problems such as: suppose $\vec{u}$ is known, and the spin half system evolves in this way for a known length of time t subject to a Hamiltonian H for which $\vec{v}$ is known but b is unknown. The state at time t now depends in a known way on the parameter b, which we could try to estimate by making a suitable measurement on the system in its final state.

2.2.3 Measurement of an observable

By a suitable interaction with the outside world one may measure an observable X. The measurement is supposed to be instantaneous. The result of the measurement is a random macroscopic change labelled by a value or outcome x while the quantum system after the measurement is in a new state corresponding to the value x. The rules of measurement are as follows:

Rule 3 (Probability of the outcome). *The probability of the outcome x is equal to $\|\Pi_{[X=x]}\psi\|^2$, the squared length of the projection of the state vector ψ onto the subspace $[X = x]$.*

Rule 4 (State after measurement). *After measurement, the state of the system is now $\Pi_{[X=x]}\psi/\|\Pi_{[X=x]}\psi\|$, the renormalised projection.*

Example 3 (Spin half). For example, in the spin half case any pure state ψ is of the form $|\vec{u}\rangle$ for some unit vector $\vec{u} \in \mathbb{R}^3$. Any observable X is of the form $a\mathbf{1} + b\vec{v}\cdot\vec{\sigma}$. This observable has eigenvalues $a \pm b$ with eigenvectors $|\pm\vec{v}\rangle$. Recall that $|\langle\vec{v} \mid \vec{u}\rangle|^2 = \frac{1}{2}(1 + \cos(\theta)) = \cos^2(\theta/2)$ where θ is the angle between $\vec{u}$ and $\vec{v}$. We compute that the probabilities of the outcomes $a \pm b$ are $\cos^2(\theta/2)$, $\sin^2(\theta/2)$ and the corresponding state of the system after the measurement is $|\pm\vec{v}\rangle$.

Example 4 (Collapse of the wave function). Consider the self-adjoint operator $X = \mathrm{diag}(1\ 2\ \ldots\ d)$, the $d \times d$ diagonal matrix with the numbers 1 to d on the diagonal. This operator has as eigenvalues these same numbers, and as eigenvectors the orthonormal basis $e_1, \ldots, e_d$ of $\mathbb{C}^d$ where the vector e_x has zeros at all positions except the x'th, where it has the value 1 (or indeed any complex number whose absolute value is 1). The eigenspace $[e_x]$ generated by the x'th eigenvector consists of all vectors whose only non-zero component is the x'th. Suppose ρ is the pure state $|\psi\rangle\langle\psi|$ corresponding to a certain normalised vector ψ, $\|\psi\|^2 = \sum |\psi_i|^2 = 1$. Measurement of X produces the values $x = 1, \ldots, d$ with probabilities

$$\begin{aligned} \mathrm{trace}(\rho\Pi_{[X=x]}) &= \mathrm{trace}(\Pi_{[\psi]}\Pi_{[e_x]}) \\ &= \mathrm{trace}(|\psi\rangle\langle\psi|\,|e_x\rangle\langle e_x|) \\ &= \mathrm{trace}(\langle e_x|\,|\psi\rangle\langle\psi|\,|e_x\rangle) \\ &= \langle e_x \mid \psi\rangle\langle\psi \mid e_x\rangle) = \psi_x\overline{\psi_x} = |\psi_x|^2. \end{aligned} \tag{2.9}$$

If the value x is found the system is now in a state with density matrix

$$\begin{aligned} \Pi_{[e_x]}|\psi\rangle\langle\psi|\Pi_{[e_x]}/|\psi_x|^2 &= |e_x\rangle\langle e_x|\,|\psi\rangle\langle\psi|\,|e_x\rangle\langle e_x|\,/|\psi_x|^2 \\ &= |\langle e_x \mid \psi\rangle|^2\,|e_x\rangle\langle e_x|\,/|\psi_x|^2 \\ &= |e_x\rangle\langle e_x|, \end{aligned} \tag{2.10}$$

i.e., in the pure state corresponding to the vector e_x.

The 'collapse of the wave function' is graphically visible here, with a transition from a state vector ψ of arbitrary complex numbers (neglecting normalisation to length one) into a vector containing zeros only, except for one position. Measurement of X forces the state vector into one of the eigenspaces $[X = x]$ and reveals the choice through the obtained value x. The probability distribution of these outcomes is given by the squared absolute values of the elements of the wave function ψ. After measurement, if the value x is found, the state now has the wave function e_x. If the same observable is measured again on the new system, the same value x will be obtained with probability one, and the state remains unaltered as the 'collapsed' state e_x.

The above example seems special but it actually gives a fairly general picture. Our observable had a special set of eigenvalues but the precise values played no role, only the fact that they were all different. Our observable was represented by a diagonal matrix, but all observables can be put into diagonal form by appropriate choice of basis. We are going to show that the picture for a general (mixed) state corresponds precisely to an ordinary probabilistic mixture of pure states. So the only really special feature of the example was that the eigenspaces of X were all one-dimensional.

2.3 Back to mixed states

Consider a quantum system with density matrix ρ. Any such density matrix can be written as $\rho = \sum p_i |i\rangle\langle i|$ for some collection of pure states $|i\rangle$. The representation is not unique. More generally one can write $\rho = \int |\psi\rangle\langle\psi| P(\mathrm{d}\psi)$ for some probability measure P over the pure state vectors ψ. We claim that the system ρ behaves indeed as an ordinary probabilistic mixture of systems $|\psi\rangle\langle\psi|$: in other words, one chooses a state ψ according to the distribution P, which thereafter undergoes measurement or Schrödinger evolution according to the rules of the previous section. For each type of evolution the result is that the system is in a new pure state, according to certain probabilities. Mixing those pure states according to the probabilities yields a density matrix describing the final, mixed state of the system. We consider separately the two kinds of evolution.

2.3.1 Schrödinger evolution of mixed states

Choose an initial state ψ_0 according to the probability distribution $P(\mathrm{d}\psi_0)$. Write $\rho(0) = \int |\psi_0\rangle\langle\psi_0| P(\mathrm{d}\psi_0)$. At time t, by (2.6) the system is then in the pure state $e^{iHt}\psi_0$. Thus with probability $P(\mathrm{d}\psi_0)$ it is in the state with density matrix $e^{iHt}|\psi_0\rangle\langle\psi_0|e^{-iHt}$. Mixing according to this distribution we find it is in the state $e^{iHt}\rho(0)e^{-iHt}$, cf. (2.8).

No surprises here, note that the Schrödinger equation for a mixed state (2.7) follows from differentiating $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$ with respect to t and then mixing with respect to $P(\mathrm{d}\psi_0)$.

The matrix $U = e^{iHt}$ has the property that $UU^* = U^*U = \mathbf{1}$, such a matrix is called *unitary*. Any unitary matrix is of the form $U = e^{iH}$ for some self-adjoint H.

2.3.2 Measurement of mixed states

Suppose we are in the pure state ψ with probability $P(\mathrm{d}\psi)$ and define $\rho = \int |\psi\rangle\langle\psi| P(\mathrm{d}\psi)$. We measure the observable X. With probability $\|\Pi_{[X=x]}\psi\|^2$ the value x is observed and the system is now in the pure state $\Pi_{[X=x]}\psi/\|\Pi_{[X=x]}\psi\|$. The joint probability of starting in state ψ and getting the outcome x is $P(\mathrm{d}\psi)\|\Pi_{[X=x]}\psi\|^2$. Let us compute the marginal probability of the outcome x, and after that the probability distribution of final state given outcome.

For the probability of the outcome x, it is equal to

$$\begin{aligned}\int_\psi P(\mathrm{d}\psi)\|\Pi_{[X=x]}\psi\|^2 &= \int_\psi \left\langle \Pi_{[X=x]}\psi \mid \Pi_{[X=x]}\psi \right\rangle P(\mathrm{d}\psi) \\ &= \int_\psi \mathrm{trace}(\left\langle \Pi_{[X=x]}\psi \right| \left| \Pi_{[X=x]}\psi \right\rangle P(\mathrm{d}\psi) \\ &= \int_\psi \mathrm{trace}(\left| \Pi_{[X=x]}\psi \right\rangle \left\langle \Pi_{[X=x]}\psi \right|) P(\mathrm{d}\psi) \\ &= \int_\psi \mathrm{trace}(\Pi_{[X=x]} \left|\psi\right\rangle \left\langle\psi\right| \Pi_{[X=x]}) P(\mathrm{d}\psi) \\ &= \mathrm{trace}(\rho\Pi_{[X=x]}).\end{aligned} \tag{2.11}$$

The conditional distribution of the final pure state $\Pi_{[X=x]}\psi/\|\Pi_{[X=x]}\psi\|$ given the outcome x is therefore $P(\mathrm{d}\psi)\|\Pi_{[X=x]}\psi\|^2/\,\mathrm{trace}(\rho\Pi_{[X=x]})$. The density matrix of this pure state is $\Pi_{[X=x]}\left|\psi\right\rangle\left\langle\psi\right|\Pi_{[X=x]}/\|\Pi_{[X=x]}\psi\|^2$. Multiplying by the conditional distribution and integrating over ψ we find the final mixed state, given the outcome x, to be $\Pi_{[X=x]}\rho\Pi_{[X=x]}/\,\mathrm{trace}(\rho\Pi_{[X=x]})$.

To summarize:

Rule 5 (Probability of the outcome). *The probability of the outcome x is equal to* $\mathrm{trace}\,\rho\Pi_{[X=x]}$.

Rule 6 (State after measurement). *After measurement, the state of the system is now* $\Pi_{[X=x]}\rho\Pi_{[X=x]}/\,\mathrm{trace}\,\rho\Pi_{[X=x]}|$, *the renormalised projection.*

2.3.3 Conclusion

In just the same way one can show that taking a system with probability distribution $P(\mathrm{d}\rho)$ in the mixed state ρ, and then either allowing it to undergo a Schrödinger evolution or measuring an observable X leads to the same final states or joint distribution of outcome and final state, as if we had started with the initial mixed state $\int \rho P(\mathrm{d}\rho)$. The rules we have given for studying a quantum system in the pure state ψ and the mixed state ρ are consistent with the picture that being given a mixed state 'is' nothing else than being given a system in a pure state with a certain probability. This should be surprising from two points of view. Firstly, different mixtures of pure states can lead to the same mixed state. Apparently there is no physical way of determining how a mixed state was prepared. For spin half, if I give you (as many times as you like) with probability half a spin up or a spin down electron, this cannot be distinguished from my giving you with probability half a spin left or a spin right electron. You will never be able to determine which procedure I am using. The second surprising fact is that the phase of the state ψ does not play any role. It is preserved during Schrödinger evolution, and in the projection of states on measurement of an observable, but it is irrelevant to the distribution of outcomes of measurement. Assuming there is no other way in which information from a quantum system becomes

available in the macroscopic real world, the phase is indeed irrelevant. Phase difference only plays a role when we are solving the Schrödinger equation for a complicated situation by considering separate solutions from sub-problems. For instance, the solution for a general initial condition might be the sum of solutions for different sets of initial conditions. In this case the phases of the different solutions do play a role as we actually add wave functions for the different solutions. An example of this phenomenon is the famous two-slit experiment where the interference pattern at the screen follows from adding the solutions to two wave equations, one each for having one slit open and the other closed.

However we indicated in chapter 1 that measurements of observables, so called simple measurements, are not the only way in which quantum systems have macroscopic interactions with the outside world. We shall next show how generalized measurements can also be brought into the picture so far through the introduction of a further key concept: product states. Later (in chapter 4.4) we will further extend the notion of measurement to continuous time rather than instantaneous measurement. But no new ideas, only mathematical complications, are involved.

2.4 Product states

If two separate quantum systems are modelled by state spaces $\mathbb{C}^{d'}$ and $\mathbb{C}^{d''}$ then it is natural to describe the system obtained when the two subsystems are brought together by the state space

$$\mathbb{C}^{d' \times d''} = \mathbb{C}^{d'} \otimes \mathbb{C}^{d''}, \tag{2.12}$$

the tensor product of the two components. It consists of all *linear combinations of* elements $\psi' \otimes \psi''$, where $\psi' \in \mathbb{C}^{d'}$ and $\psi'' \in \mathbb{C}^{d''}$. If ψ' and ψ'' are themselves linear combinations, then one can expand bilinearly to express $\psi' \otimes \psi''$ in terms of tensor products of the components. An orthonormal basis of $\mathbb{C}^{d' \times d''}$ is formed by the tensor products of elements of bases of $\mathbb{C}^{d'}$ and $\mathbb{C}^{d''}$. The dimension of the product space is indeed the product of the dimensions of its components. The idea behind this construction is that a) any combination of states of the two subsystems must be a state of the joint system, and b) because the joint system is also a quantum system, arbitrary linear combinations of states are also states.

One can also now form tensor products of operators, say $X = X' \otimes X''$, defined by linearity and the rule $(X' \otimes X'')(\psi' \otimes \psi'') = X'(\psi') \otimes X''(\psi'')$. In particular, if ρ' and ρ'' are mixed states on the component systems, then $\rho = \rho' \otimes \rho''$ is a mixed state on the compound system. Similary if ψ' and

ψ'' are pure states of the components, then $\psi = \psi' \otimes \psi''$ is a pure state of the compound system. However it is important to realise that the tensor product space allows many other pure states, many other mixed states, and many other observables, than these special cases!

Considering mixed states as mixtures of pure states, the state $\rho = \rho' \otimes \rho''$ is the same as the mixture of states $\psi' \otimes \psi''$ where the components are chosen *independently* according to the mixing distributions involved in ρ' and ρ''.

An observable X' on the first subsystem can be represented by $X' \otimes \mathbf{1}''$ on the joint system, and similarly for an observable X'' on the second subsystem. Note that $X' \otimes \mathbf{1}''$ and $\mathbf{1}' \otimes X''$ commute, so separate measurements on the two subsystems are automatically compatible in this construction. The joint probability distribution of joint measurements of the two just described observables, when the joint system is in the product state $\rho = \rho' \otimes \rho''$, is that of independent measurements of X' on ρ' and X'' on ρ'', as we will see in the next section. The fun starts when we realise that there are also states in the product space which are not tensor products of separate states on the two subsystems. Pure states which are linear combinations of more than one product state $\psi' \otimes \psi''$ are called *entangled*.

We will come back to examples after we have discussed the joint measurement of several observables in the next section.

2.4.1 Marginalization of a product state

Consider a quantum state ρ of a product-system. If one only considers measurements on the first component of the system, then one does not need to know ρ in its entirity. It is not difficult to see that all predictions about the first component are the same as for a one-component system whose state is the *partial trace, over the second system* of the joint state. Written out componentwise, $(\mathrm{trace}_2\, \rho)_{i',j'} = \sum_{i''} \rho_{i',i'';j',i''}$.

2.5 Repeated measurements

2.5.1 Repeated and joint measurements

Having described the effects of a measurement of a single observable, we can study the effects of repeated measurements of a sequence of observables. We simply iterate rules 5 and 6, if we prefer a description in terms of density matrices (general mixed states), or rules 3 and 4, if we prefer to work with wave functions (pure states). The two sets of rules are mutually consistent using the fact that a probabilistic mixture of states is represented by the same

mixture of corresponding density matrices. We do not mix state vectors, but mix projectors onto the subspaces generated by state vectors.

It follows immediately from either representation that repeated measurement of the same observable simply repeats the first obtained value, and does not alter the state further. The first measurement projects the state into an eigenstate of the operator, and on repetition, it stays there. If different observables are measured in sequence, the results are strikingly different according to whether the corresponding operators commute or not.

Commuting self-adjoint operators represent so-called compatible observables, that is, physical quantities which in principle can be measured simultaneously. We describe the mathematical facts behind this statement which at least allow it to make sense. Suppose X and Y are commuting self-adjoint operators. Recall that this implies that they can both be written as a function of a third operator Z, say $X = f(Z)$, $Y = g(Z)$. Of course Z can be replaced by any one-to-one function of Z. Subject to this non-uniqueness, we choose Z 'minimally', to be precise, with as few as possible eigenspaces. The fact that X and Y commute implies that the subspaces $[X = x, Y = y] = [X = x] \cap [Y = y]$, as x and y vary through the spectra of X and Y, form a decomposition of $\mathbb{C}^d$ (orthogonal subspaces spanning $\mathbb{C}^d$). We can then specify Z by letting its eigenspaces be the nontrivial subspaces $[X = x, Y = y]$ (i.e, those for which $[X = x] \cap [Y = y] \neq \{0\}$). We assign each of these a distinct value z.

Joint measurement of X and Y could conceivably be modelled now in three different ways: 1) first measure X, then measure Y; 2) first measure Y, then measure X; 3) measure Z and report the values $x = f(z)$ and $y = g(z)$. The reader should verify that the specifications are probabilistically identical; i.e., they result in a pair of values (x, y) and a new state $\rho_{x,y}$ depending on the original state ρ where the joint distribution of values x, y does not depend on which of the three routes we took, and the final state $\rho_{x,y}$ also depends only on x, y and ρ and not on the route which was taken.

Whichever route the reader prefers, the answer he or she obtains should be the following natural generalization of rules 5 and 6: the joint measurement of commuting observables X and Y on a quantum system in state ρ produces the pair of values x, y and leaves the system in the state

$$\rho_{x,y} = \frac{\Pi_{[X=x,Y=y]}\rho\Pi_{[X=x,Y=y]}}{\mathrm{trace}(\rho\Pi_{[X=x,Y=y]})} \tag{2.13}$$

with probability

$$\mathrm{trace}(\rho\Pi_{[X=x,Y=y]}). \tag{2.14}$$

The description in terms of pure states ψ is similarly perfectly analogous to rules 3 and 4.

It is easy to check that the marginal distribution of the measurement of X in a joint measurement of compatible X and Y is the same as that of the measurement of X alone. The experimental set-ups would generally be quite different since the resulting possible states of the system after the joint measurement are not the same as after the single measurement: the decomposition into subspaces $[X = x, Y = y]$ of $\mathbb{C}^d$ is finer than the decomposition into subspaces $[X = x]$, unless Y is actually a function of X.

If X and Y do not commute, and one is measured after the other, then the distribution of values of each observable depends on the order in which they are measured. The possible final states of the system also depend on the order.

2.5.2 The unconscious quantum statistician

The above discussions can be extended, obviously, to the joint observation of any number of compatible observables. Let us summarize one of the main conclusions as a formal rule:

Rule 7 (Marginal distributions). *If one set of observables is a subset of a larger set of compatible observables, then the joint distributions of the variables in the subset when they alone are measured, and when the whole set is measured, are the same.*

Now suppose that X and Y are two commuting observables, and $Z = X + Y$ is their sum. Z commutes with X and Y and it makes sense to consider a joint measurement of all three variables. By the natural extension of (2.14) above, the probability of obtaining any triple of values (x, y, z) is given by

$$\text{trace}(\rho \Pi_{[X=x, Y=y, Z=z]}). \tag{2.15}$$

The reader should verify that this probability distribution is the same as the distribution obtained by measuring just X and Y simultaneously, yielding pairs of values (x, y), and then appending $z = x + y$ to generate a triple (x, y, z). More generally:

Rule 8 (Law of the unconscious quantum statistician). *If a number of compatible observables satisfy a functional constraint $f(X, Y, Z, \dots) = \mathbf{0}$ (e.g., $X + Y - Z = \mathbf{0}$), then the values obtained in a joint measurement of the same observables will satisfy the same functional constraint with probability one.*

These results are not difficult to establish, at least in the discrete case, but they are hardly ever stated explicitly in treatments of quantum probability. We feel they deserve to brought into the open. They are perfectly analogous to the 'law of the unconscious statistician' known from ordinary probability theory, according to which one can compute the expectation of a function of a random variable directly by integration with respect to the distribution of the original variable, rather than via a computation of the distribution of the transformed variable.

2.5.3 Conditioning

Measurement of an observable X on a system in state ρ produces the value x with probability $\mathrm{trace}(\rho\Pi_{[X=x]})$; the system has then made a transition into the new state $\Pi_{[X=x]}\rho\Pi_{[X=x]}/\,\mathrm{trace}(\rho\Pi_{[X=x]})$. Suppose we now go on to measure a new observable Y. Denote by $X_\rho, Y_{\rho:X}$ a pair of random variables having the joint distribution of values found when first X is measured, then Y. The previously established rules tell us

$$\mathrm{E}(Y_{\rho:X}|X_\rho = x) \;=\; \frac{\mathrm{trace}(\Pi_{[X=x]}\rho\Pi_{[X=x]}Y)}{\mathrm{trace}(\rho\Pi_{[X=x]})}. \tag{2.16}$$

We consider this as a function of x, and hence also of the operator X and the random variable X_ρ. Since by definition $f(X) = \sum f(x)\Pi_{[X=x]}$ we find for the function of the operator, which we denote by $\mathrm{E}_\rho(Y|X)$,

$$\mathrm{E}_\rho(Y|X) \;=\; \sum_x \frac{\mathrm{trace}(\Pi_{[X=x]}\rho\Pi_{[X=x]}Y)}{\mathrm{trace}(\rho\Pi_{[X=x]})}\Pi_{[X=x]}. \tag{2.17}$$

The expected value of a measurement of this observable is

$$\mathrm{E}\big((\mathrm{E}_\rho(Y|X))_\rho\big) \;=\; \mathrm{trace}\big(\rho\mathrm{E}_\rho(Y|X)\big) \;=\; \sum_x \mathrm{trace}(\Pi_{[X=x]}\rho\Pi_{[X=x]}Y) \tag{2.18}$$

which is typically unequal to $\mathrm{E}(Y_\rho) = \mathrm{trace}(\rho Y)$. It is the same a) when Y commutes with X, b) when X commutes with ρ. The same conditions guarantee the more general result that the random variables Y_ρ and $Y_{\rho:X}$ have the same distributions, since

$$\Pr\{Y_{\rho:X} = y|X_\rho = x\} \;=\; \frac{\mathrm{trace}(\Pi_{[X=x]}\rho\Pi_{[X=x]}\Pi_{[Y=y]})}{\mathrm{trace}(\rho\Pi_{[X=x]})} \tag{2.19}$$

yielding

$$\Pr\{Y_{\rho:X} = y\} \;=\; \sum_x \mathrm{trace}(\Pi_{[X=x]}\rho\Pi_{[X=x]}\Pi_{[Y=y]}). \tag{2.20}$$

Condition a) says X and Y are both functions of a third observable, say Z, and can be measured simultaneously, with the marginal distributions of their measurements remaining unchanged according to the law of the unconscious quantum statistician. Condition b) says that the quantum system can be considered to be in an ordinary random mixture of pure states in each of which X has a fixed value. Measuring X does not change the state of the system, hence it does not change the probability distribution of a subsequent measurement of the observable Y.

This account of conditioning shows that it is easy, at least in the discrete case, to introduce conditioning and joint distributions in a physically meaningful way into quantum probability. The fact that the distributions and in particular expectations of Y_ρ and $Y_{\rho:X}$ are generally different, is a natural consequence of the fact that measuring an observable changes the quantum system.

Modern quantum probability theory employs a notion of conditioning where $\mathrm{E}(Y|X)$ is defined to be that function of X, say $g(X)$, such that $\mathrm{trace}(\rho(Y - g(X))^2) = \mathrm{E}(Y - g(X))^2_\rho$ is minimal. This imitates a defining property of classical conditional expectation, and has such nice properties as $\mathrm{trace}(\rho\mathrm{E}(Y|X)) = \mathrm{trace}(\rho Y)$, i.e., $\mathrm{E}(\mathrm{E}(Y|X)_\rho) = \mathrm{E}(Y_\rho)$. However, if Y and X do not commute, then neither do Y and $g(X)$, and hence the observable $Y - g(X)$ has no physical interpretation in terms of the incompatible observables X and Y.

2.6 Quantum randomized measurement

2.6.1 Quantum randomization

We now have a background of theory and examples in quantum probability, from which we can start to study problems in quantum statistics. For instance, given a particle is in an unknown state $\rho = \rho(\theta)$ parametrised by a quantity θ of physical significance, how can we estimate θ? The obvious answer is: take an observable X and measure it. The probability distribution of the result of measurement depends on θ and we now have a classical inference problem. Perhaps we can repeat this experiment n times, each time independently producing a particle in the same state $\rho(\theta)$ and measuring the same observable. Now we have a classical i.i.d. sample from the distribution of measurements of X on a particle of state $\rho(\theta)$.

This description suggests that the main element of choice is the observable X and the sample size n. However, the theory we have built up suggests many more possibilities.

For instance, measure one observable, forcing a collapse of the wave function, and record the value obtained; then measure another; finally take some numerical function of both the values. Do the same thing in any number of stages, not just two.

Between measurements of basic observables allow the quantum system to evolve unitarily under various Hamiltonians H.

Allowing a time evolution and then measuring an observable is actually equivalent to measuring a different observable; use the fact that

$$\text{trace}\big((e^{-iHt}\rho e^{iHt})\,A\big) \;=\; \text{trace}\big(\rho\,(e^{iHt}Ae^{-iHt})\big). \tag{2.21}$$

This transferral of the unitary operator e^{iHt} between state and observable is called the switch from the Schrödinger picture of quantum mechanics, according to which states evolve but observables remain fixed, to the Heisenberg picture, in which it is the other way round. In practice, measurements of observables are often realised this way: measure with, say, a single physical apparatus, e.g., the same Stern-Gerlach device, but first subject the particles to time evolution according to various Hamiltonians H.

Yet more possibilities come up if we bring randomisation, in the classical sense, into the picture. Choose observables or Hamiltonians at random according to the results of classical randomisation devices like throwing dice, tossing coins, picking cards from a shuffled pack, or pulling balls out of vases. Let the measurements and the measurement devices be macroscopically perturbed by macroscopic random noise. Suppose the lengths of time of Hamiltonian evolutions are not exactly known but are random.

A fundamental new option is provided by our mathematical construction of joint quantum systems from separate components. We call this construction *quantum randomisation*. Suppose we want to study a particle in a state $\rho = \rho(\theta)$. Take a separate quantum system with another, possibly much larger state space, in a known state ρ'. Allow the particle whose state ρ we are interested in to interact with the second system. The two systems taken together are modelled by the state $\tilde{\rho} = \rho \otimes \rho'$ in the product space. Measure an observable $\tilde{X}$ on the compound system. This could for instance be realised by letting the two particles interact together for some time under some Hamiltonian $\tilde{H}$ (not of the special form $H \otimes \mathbf{1}' + \mathbf{1} \otimes H'$) and then measuring an observable on the second subsystem.

The point of the parenthetical remark is that for a Hamiltonian of that special kind, the energy of the joint system is the sum of the energies of the

separate systems, hence no interaction takes place; the two subsystems would evolve independently according to their own Hamiltonians. However if there really is interaction between the two particles, what we generate in this way cannot be represented as the measurement of an ordinary observable on the first system on its own.

After this measurement one could separate the two systems again; or to put it another way, only consider further operations on the first system. Our rule for the state of a quantum system after measurement of an observable followed by the operation of taking the partial trace can be used to deduce the state of the original subsystem after a specified interaction and measurement. Now yet more measurements can be made on the system under study.

Quantum randomisation need not just be thought of as based on interaction between *different particles*. The tensor product construction can also be used to represent different aspects of the same particle; e.g., spin and location; or (anticipating the extension to continuous observables) location in the x direction and location in the y direction; and so on. For instance in a Stern-Gerlach apparatus we measure spin by observing location.

By the way, if we are allowed to make n measurements on n independently and identically prepared particles, we might also be allowed to make one measurement on the compound system in the state $\rho^{\otimes n} = \rho \otimes \cdots \otimes \rho$ obtained by considering these n particles as one system. We come back to this later, but for the time being restrict attention to measurement of just one particle.

The practical possibilities sketched above are vast; certainly when one realises that one can combine all these operations in sequence. From a mathematical point of view however, it fortunately turns out that all the measurement strategies we have mentioned, and indeed any we may have omitted, can be brought under a single heading called 'generalised observables'. Moreover, at least in principle, every generalised observable can be realised by a quantum randomisation. In the next subsection we take an abstract look at these matters. The key point is that all the measurement procedures we have discussed are *affine*, that is, commute under the taking of mixtures. In plain probabilistic language, measurement on a mixed state is equivalent to mixing the results of measurement on the components of the mixture. This follows as far as time evolution is concerned from the linearity of the Schrödinger equation, as far as measurement of observables is concerned from the linearity in both the distribution of measured values (linearity of trace operator) and in the resulting unnormalised states (projections), as far as classical randomisation is concerned it follows from the linearity of convolution, and as far as quantum randomisation is concerned it follows from the linearity of the tensor product in each component and again the linearity of the trace operator.

2.6.2 Generalized measurements

First we repeat some of the material of chapter 1 by mathematically extending the notion of observable. We do this in as simple way as possible, while preserving the mixing properties of measurement as far as the distribution of the measured values is concerned.

Recall that an observable X stands in one-to-one correspondence with a collection of values and orthogonal subspaces. Replacing the subspaces by the projectors onto those subspaces, and considering the values of the observable merely as arbitrary labels, one can identify an observable with a *orthogonal resolution of the identity* $A_1, \ldots, A_n$

$$A_1 \ + \ \ldots \ + \ A_n \ = \ \mathbf{1} \tag{2.22}$$

where the A_i are projectors (idempotent self-adjoint non-negative operators), mutually orthogonal and commuting: $A_i A_j = A_j A_i = \mathbf{0}$ for $i \neq j$; $A_i^2 = A_i$; $\psi^* A_i \psi \geq 0$ for all ψ. Here, n cannot exceed the dimension d of the space $\mathbb{C}^d$ if the A_i are all nontrivial.

We call *any* collection of non-negative self-adjoint operators A_i satisfying $\sum_i A_i = \mathbf{1}$ a (generalised or non-orthogonal or arbitrary) resolution of the identity. Now the number of components n is arbitrary. It can be verified (see problem 3) that the A_i are all idempotent ($A_i^2 = A_i$) if and only if they are *orthogonal projectors*. Thinking geometrically, an idempotent self-adjoint operator is a projector, and a sum of projectors can only be the identity if the projectors project onto a decomposition of the Hilbert space.

Let ρ be any density matrix. It is a mixture of pure states $\rho_\psi = \psi\psi^* = \Pi_{[\psi]}$, $\rho = \int \Pi_{[\psi]} P(\mathrm{d}\psi)$. Now let (A_i) be a resolution of the identity, and define $p_i = \mathrm{trace}(\rho A_i)$. Since the A_i are nonnegative, we have

$$p_i = \int \mathrm{trace}(\Pi_{[\psi]} A_i) P(\mathrm{d}\psi) = \int \psi^* A_i \psi \, P(\mathrm{d}\psi) \geq 0. \tag{2.23}$$

Since $\sum A_i = \mathbf{1}$ we have

$$\sum p_i = \sum \mathrm{trace}(\rho A_i) = \mathrm{trace}(\rho(\sum A_i)) = \mathrm{trace}(\rho) = 1. \tag{2.24}$$

Thus the p_i form a probability distribution.

For a given resolution of the identity (A_i) we consider the mapping

$$\rho \ \mapsto \ (p_i = \mathrm{trace}(\rho A_i) : i = 1, \ldots, n) \tag{2.25}$$

taking density matrices into probability distributions on $\{1, \ldots, n\}$, i.e., into the n-dimensional simplex. This mapping is *affine*, in other words, it commutes with the taking of convex combinations. Note that both the space

of density matrices and the n-dimensional simplex are convex sets though of essentially different character. The simplex looks like a tetrahedron, with the degenerate distributions forming the corners. The space of density matrices has a smooth boundary (all density matrices of less than full rank), with the pure states (rank one density matrices) forming a continuum on part of the boundary. They supply the extreme points. As we saw, if $d = 2$ the space of density matrices can be affinely identified with the three dimensional unit ball, the so-called Poincaré representation; the extreme points now form all of the boundary. Any density matrix is the convex combination of a finite number of pure states (the same number as the rank of the matrix) but the representation is not unique and a continuum of pure states is required to generate all of the density matrices. The simplex has a finite number of extreme points and every point in the simplex is a unique convex combination of them.

It is an easy theorem that *every* affine map from the density matrices to the simplex can be expressed as $\rho \mapsto (p_i = \mathrm{trace}(\rho A_i) : i = 1, \ldots, n)$ for some resolution of the identity (A_i). The proof runs as follows: consider a single p_i as real function of ρ; extend to all self-adjoint operators by linearity; it has to be of the form $\mathrm{trace}(\rho A_i)$ for some self-adjoint operator A_i; then use the fact that p_i forms a probability distribution to get the remaining required properties of nonnegativity and adding to $\mathbf{1}$.

Thus all the measurement strategies we described in the last section correspond to resolutions of the identity. In particular this applies to *quantum randomisation*—measurement of a classical observable on a compound system formed by taking the tensor product of the system of interest with another system.

A celebrated theorem of Naimark (called the Gel'fand-Naimark-Segall construction in functional analysis), states that *quantum randomisation exhausts the class of all possible measurements.* We have to show that for any resolution of the identity A_i in $\mathbb{C}^d$, there exist d',ρ' and an orthogonal resolution of the identity $\tilde{A}_i$ in $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ such that

$$\mathrm{trace}(\rho \otimes \rho' \tilde{A}_i) = \mathrm{trace}(\rho A_i) \tag{2.26}$$

for all states ρ on $\mathcal{H}$. A proof of the theorem in this discrete setting is given in Holevo (1982, ...) and is simple and explicit. One says that $(\tilde{A}_i)$ forms a *realization* of (A_i).

`proof to be included!`

It can be shown that this measurement can be implemented by first allowing a Hamiltonian evolution of the joint system, and then measuring an observable on the ancillary system; thus $\tilde{A}_i = U\Pi_{[X=i]}$ for some unitary U on

the joint system and an observable X which only operates on the auxiliary system.

More generally one can generate measurements taking a continuum of different values by allowing the ancillary system to be infinite dimensional. The results we have just mentioned generalise: the outcome of any measurement is representable by an operator valued probability measure, and any operator valued probability measure is representable by the measurement of an observable on a compound system.

We described not just the distribution of results of measurement of ordinary observables, but also the new state of the system after measurement. One may ask if there is a similar theory for generalised observables. If the generalised observable was indeed realised in a quantum randomisation experiment then we know in principle the answer to this question. But clearly, there can be many different quantum randomizations which reproduce the same OProM, but lead to quite different final states. Is there a simple way to describe the possible final states resulting from this procedure? And is it possible to transform a quantum state in other ways than by quantum randomization?

The answer is given by another celebrated theorem: the Kraus representation theorem; see Kraus (1983), Ozawa (1984), Preskill (1997), Werner (1997), Holevo (1999); it is based on a result from functional analysis called the Stinespring theorem. It shows that *there are no other ways to transform a quantum state than by quantum randomization.* Moreover, the theorem gives a simple representation of all physically possible mappings from input states to output states. Consider a measurement with discrete outcome space, which can produce the outcome x with probability p_x, in which case the quantum system is transformed into state ρ_x. Define $\widetilde{\rho}_x = p_x\rho_x$ and note that the mapping from ρ to $\widetilde{\rho}_x$ is linear, and sends density matrices to unnormalised density matrices with trace less than one. Moreover, from $\widetilde{\rho}_x$ one can recover $p_x$and ρ_x. Suppose we place another quantum system ρ_0 alongside of the state ρ but do not let it interact in any way with the original system and the measurement process, then our measurement transforms the two systems together from $\rho \otimes \rho_0$ to $\widetilde{\rho}_x \otimes \rho_0$, when the outcome x is observed (and does this with the same probability as before). We are now looking at an extended measurement on product systems, and we know how it operates on all tensor products. It must be linear by the interpretation of mixed states as statistical mixtures. By linearity and by its action on product states, its action on all states in the product system is determined; and this must be an action which sends density matrices to unnormalized density matrices (with trace less than one). It turns out that this requirement, termed *total positivity*, restricts the original mapping to have the form: $\rho \mapsto \widetilde{\rho}_x = \sum_y V^*_{x,y}\rho V_{x,y}$ where the $V_{x,y}$ (a

countable collection of matrices) must just satisfy $\sum_y V_{x,y} V_{x,y}^* \leq 1$.

Moreover, any such mapping can be represented can be represented as a quantum randomization: there exists an auxiliary system in the state ρ', a unitary matrix U on the joint system, and two commuting observables X and Y on the joint system, such that $V_{x,y} \rho V_{x,y}^* = \mathrm{trace}_2(\Pi_{[X=x,Y=y]} U(\rho \otimes \rho') U^* \Pi_{[X=x,Y=y]})$. This means that the measurement procedure can be interpreted as follows. The system ρ is brought into interaction with another system ρ'. A unitary evolution transforms the two initially independent systems into one joint system. An observable $Z = (X, Y)$ of the auxiliary system is measured, projecting the joint state to an eigenstate. The outcome of Y is not observed, but that of X is, so that the final state of the original system is the partial trace of the mixture over possible outcomes y, of the state resulting when the outcome is (x, y).

If the outcome x of the measurement is not discrete, the description of the transformation of the input state to the output state is mathematically rather delicate and involves subtle functional analytic considerations, but the picture remains essentially unchanged, especially when we restrict attention to finite dimensional input systems and measurements taking values in Borel subsets of the real line (or on measureable spaces which are measure-theoretically isomorphic to a Borel subset of the real line, i.e., Borel subsets of separable metric spaces. See Ozawa (1984, 1985, 1988), Holevo (1999).

A considerable modelling effort is required, for a specific measurement situation, to decide how all the operators involved are to be specified. This involves doing a lot of real physics in the grey area between quantum (microscopic) and classical (macroscopic), involving heuristics, approximations, and hard calculations.

We shall return to transformations of state in the chapter on quantum stochastic processes. There we shall study measurement continuous in time. When we study for instance radioactive decay, observations are typically random time points of definite events (the clicking of a Geiger counter), rather than random events observed at one definite time point. This type of situation is modelled with a product space to represent the combined system of atom and emitted particle; the observation process is modelled by measurements,closely repeated in time, to detect the emitted particle.

2.7 Ordinary versus quantum probability

Over the years, a hotly debated issue has been whether the randomness of quantum mechanics can be described using ordinary probability theory. Of course that strongly depends on what one means by 'ordinary'. Physicists

on the one hand, and pure mathematicians on the other, have generally agreed that quantum probability is a different kind of probability (Feynman, 1953) and that the axioms of Kolmogorov are too restrictive (Accardi, ...; Kümmerer and Maassen, 1996). Again, whether one holds this view or not depends very much on what is considered to be the ambit of probability theory. We have argued that, in any case, the randomness in quantum mechanics is definitely the business of ordinary probabilists and statisticians.

According to most probabilists, Hilbert's tenth problem, to axiomatize parts of physics and in particular probability theory, was solved by Kolmogorov (1933) with his now familiar axiomatisation. The germinating event was the discovery of the Radon-Nikodym theorem, allowing conditioning to be given a firm measure-theoretic basis (the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ had been available for some decades). However a year before the appearance of Kolmogorov's work, von Neumann (1932) gave an axiomatization of quantum mechanics nowadays called *quantum probability*, and seen by many as a real alternative to classical or Kolmogorov probability. In a mathematical sense it is a weaker set of axioms so is applicable to a wider class of real-world phenomena. In this axiomatization, the probability space Ω with its elements, the elementary outcomes ω, seem to be missing. But one can recognise *analogues* of random variables X, expectations, even probabilities of events (expectations of 0–1 valued random variables) and so on.

The analogy can be useful. For instance, the notion of 'observable' plays a similar and equally fundamental role in quantum probability as that of 'random variable' in ordinary probability. Our notation has emphasized the analogy, for instance we use the symbol X to denote a typical observable even though it is represented by a mathematical object quite different from that used to represent a random variable. However the analogy also has its dangers and in the past has caused plenty of confusion.

'Quantum probability' is the branch of mathematics where we study the abstract structure of observables X and density matrices ρ, with expectation given by $\mathrm{trace}(\rho X)$. This mathematical structure can be seen as a non-commutative extension of the Kolmogorov structure of random variables and probability measures. However as we have described it above, there is no conflict whatsoever between these different models of probability theory. Take a quantum system, choose a set of compatible observables, and measure them: the rules above tell us the probability distribution—in the classical sense—of the results obtained, including the resulting state of the system. 'Mixed states' are just probabilistic mixtures. If one wants to go to the trouble of explicitly constructing a probability space for the results of a specified experiment, there is no barrier to doing this. Conditioning can be effortlessly introduced into the picture—the distribution of a measurement of Y, after

a measurement of X has been taken yielding the value x, can be calculated straightforwardly. If X and Y commute this will be the same as the usual conditional distribution of the Y measurement given the X measurement computed from the joint distribution of a simultaneous measurement of the two (and it is only for commuting variables that simultaneous measurement is physically speaking meaningful).

Chapter 3

Infinite dimensional systems

3.1 Introduction

After some introductory comments, we will give in Section 3.2 a survey of the necessary operator theory for infinite dimensional Hilbert spaces. The survey is of necessity both technical and lengthy but we have tried to keep it as compact as possible. A quite readable source for this material is Chapters VI–VIII of Reed and Simon (1972), *Methods of Mathematical Physics*, Part I. Still, they go into many difficult topics which we do not need to know about while they skip very briefly over what for us are some of the key issues.

In the third section of the chapter, we show how the theory can be used to represent states and observables as in the discrete case, and conclude with a summary of the key results which emphasizes the close resemblance to the discrete case. The reader might prefer to read these sections in reverse order.

For interesting and important applications as well as for mathematical elegance we need to develop an infinite-dimensional theory. Naturally enough, if we replace the finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$ by an infinite-dimensional space $\mathcal{H}$, but with a countable basis, much of the previous theory can be easily extended just by replacing finite sums everywhere by infinite sums. Surprisingly perhaps, this simple extension not only yields discrete probability models with countable numbers of different outcomes but also provides continuous models. This does depend on some subtle mathematical issues. We will continue to identify observables ('physical quantities') with self-adjoint operators, but in doing so we will extend the notion of operator a little, including so-called unbounded operators which are linear maps defined on a dense subset of the Hilbert space $\mathcal{H}$ rather than the whole space. However, as we have emphasized, it is the accompanying projector-valued probability measure which is the most important feature of an observable.

The operator representation is just a convenient packaging of these items. The key mathematical result which allows this correspondence to continue to hold is the celebrated *spectral theorem*, which establishes a one-to-one correspondence between unbounded self-adjoint operators and the *projector-valued probability measures* on the real line. We will also be able to include generalised observables in the picture through the concept of *operator-valued probability measures*. Recall that these notions were introduced in definitions 6 and 7 in Chapter 1. Those definitions remain valid, simply replacing the words matrix by operator and the space $\mathbb{C}^d$ by $\mathcal{H}$.

3.2 Unbounded operators.

Let $\mathcal{H}$ be a *separable* complex Hilbert space. It consequently has a countable orthonormal basis (in fact, many different bases). An operator X on $\mathcal{H}$ is a continuous linear map from $\mathcal{H}$, or just from a dense subset of $\mathcal{H}$, to itself. If it is defined everywhere it is automatically a bounded operator, i.e., $\sup_{\psi:\|\psi\|=1} \|X\psi\| < \infty$; if it cannot be defined (so as to be continuous) everywhere it is called unbounded and correspondingly $\sup_{\psi:\|\psi\|=1} \|X\psi\| = \infty$ where the supremum is now taken over ψ in the domain of X.

It is useful to know that in complex Hilbert space an operator can be defined by just specifying the values of $\langle\psi|X\psi\rangle$ for all ψ, or all ψ in a dense subset of $\mathcal{H}$. We do not need to know *how* to do this, but as a hint for the interested reader we mention that it depends on the so-called polarization identity

$$\begin{aligned}\langle\phi|X\psi\rangle \;=\; \tfrac{1}{4}\Big(&\langle(\phi+\psi) \;|\; X(\phi+\psi)\rangle \\ &-\langle(\phi-\psi) \;|\; X(\phi-\psi)\rangle \\ &+i\langle(\phi+i\psi) \;|\; X(\phi+i\psi)\rangle \\ &-i\langle(\phi-i\psi) \;|\; X(\phi-i\psi)\rangle \Big)\end{aligned} \tag{3.1}$$

which allows us to express $\langle\phi|X\psi\rangle$ in terms of what has been given. After that, if we know the inner product of ϕ with $X\psi$ for every ϕ (or for every ϕ in a dense subset of $\mathcal{H}$), we know the value of $X\psi$.

The adjoint of an operator X is the operator X^* defined by requiring

$$\langle X^*\phi|\psi\rangle \;=\; \langle\phi|X\psi\rangle \tag{3.2}$$

for all ψ in $\mathrm{dom}(X)$. Its (we emphasize, maximal) domain consists of precisely those ϕ for which the linear function $\langle\phi|X\psi\rangle$ of $\psi \in \mathrm{dom}(X)$ can be extended continuously to all of $\mathcal{H}$. If so, this continuous linear function must by the

Riesz representation theorem (the self-duality of Hilbert space) be an inner product $\langle\eta|\psi\rangle$ for some η in $\mathcal{H}$, which is then the value of $X^*\phi$. An operator is called self-adjoint if $X^* = X$, by which we imply that the two functions also have the same domains. Some authors use the word 'Hermitian' instead of 'self-adjoint', but for others 'Hermitian' does not imply equality of the domains of X and X^*, so that being Hermitian is a weaker property of an operator than being self-adjoint.

A projector on $\mathcal{H}$ is a bounded self-adjoint idempotent operator. Its image is a closed subspace of $\mathcal{H}$ and the operator does simply project onto this subspace. The self-adjoint operators stand in one-to-one correspondence with the *projector-valued probability measures*. A projector-valued probability measure or ProProM is a mapping $B \mapsto \Pi(B)$ from the Borel sets of the real line to the projectors on $\mathcal{H}$ such that

$$\Pi(\emptyset) = \mathbf{0}, \tag{3.3}$$

$$\Pi(\mathbb{R}) = \mathbf{1}, \tag{3.4}$$

and for every countable collection of *disjoint* Borel sets B_i

$$\sum_i \Pi(B_i) = \Pi(\bigcup_i B_i). \tag{3.5}$$

The countable sum of operators in the property of σ-additivity is defined as the strong operator limit of its partial sums; i.e., by requiring $\sum_{i=1}^{\infty} \Pi(B_i)\psi = \lim_{n\to\infty} \sum_{i=1}^{n} \Pi(B_i)\psi$ for all $\psi \in \mathcal{H}$. Correspondingly the closed subspaces onto which the projectors project are related as follows: the *closure* of the orthosum of the images $\Pi(B_i)(\mathcal{H})$ equals the image of the sum of the projectors $\Pi(\cup B_i)(\mathcal{H})$.

As we said there is a one-to-one correspondence between self-adjoint operators and projector-valued (probability) measures. We will now show how one direction of this correspondence works. Let a projector-valued measure $(B \mapsto \Pi(B))$ be given. Anticipating the result of the construction, write $[X \in B]$ for the closed subspace onto which $\Pi(B)$ projects, and similarly write $\Pi_{[X\in B]} = \Pi(B)$. The operator X will be determined by specifying $\langle\psi|X\psi\rangle$ for ψ in an, as large as possible, dense subset of $\mathcal{H}$. This subset will also be the domain of the operator. Here is the recipe:

$$\langle\psi|X\psi\rangle \;=\; \int x\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle \tag{3.6}$$

for all ψ such that

$$\int x^2\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle \;<\; \infty, \tag{3.7}$$

where the notation $\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle$ denotes ordinary integration with respect to the real measure

$$\begin{aligned}\mu(B) &= \langle\psi|\Pi_{[X\in B]}\psi\rangle \\ &= \langle\psi|(\Pi_{[X\in B]})^2\psi\rangle \\ &= \langle\Pi_{[X\in B]}\psi|\Pi_{[X\in B]}\psi\rangle \\ &= \|\Pi_{[X\in B]}\psi\|^2.\end{aligned} \tag{3.8}$$

Here we exploited the self-adjointness and idempotence of a projector to write the measure in such a form that it is evident that it is a bounded, nonnegative, real measure, of total mass $\|\psi\|^2$. We have $\|X\psi\|^2 = \int x^2\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle$ for ψ in the domain of X.

Borel functions of self-adjoint operators can be defined via their projector-valued measures as follows: $f(X)$ is defined by

$$\langle\psi|f(X)\psi\rangle \;=\; \int f(x)\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle \tag{3.9}$$

for all ψ (forming the domain of $f(X)$) such that

$$\int f(x)^2\langle\psi|\Pi_{[X\in \mathrm{d}x]}\psi\rangle \;<\; \infty. \tag{3.10}$$

The domain is dense in $\mathcal{H}$, as can be seen by noting that it contains the union of the closed subspaces $[X \in f^{-1}([-n,n])]$. This set may not be all of $\mathcal{H}$ but it can be argued to be dense by using the fact that $\Pi_{[X\in f^{-1}([-n,n])]} \to \mathbf{1}$ as $n \to \infty$ (strong operator limit).

In fact the projector-valued measures of X and $f(X)$ are related by the fact that $f(X)$ has the projector-valued measure

$$B \;\mapsto\; \Pi_{[X\in f^{-1}(B)]}. \tag{3.11}$$

In view of the one-to-one correspondence between self-adjoint operators and projector-valued measures, this supplies an alternative way to define $f(X)$. In particular we obtain for the indicator function 1_B that

$$1_B(X) \;=\; \Pi_{[X\in B]}. \tag{3.12}$$

One can also define $f(X)$ as the measure-theoretic integral of the real function f with respect to the projector-valued measure $B \mapsto \Pi_{[X\in B]}$,

$$f(X) \;=\; \int f(x)\Pi_{[X\in \mathrm{d}x]}. \tag{3.13}$$

This definition follows the usual route in defining the Lebesgue integral (via simple functions and nonnegative functions), taking special care with domain questions of the resulting sequences of operators. In the literature usually integrals of bounded functions are defined in this way, and then unbounded functions treated by the usual approach in Riemann integration, i.e., by taking the limit as $n \to \infty$ of integrals over $[-n, n]$.

The smallest closed set $A \subseteq \mathbb{R}$ such that $[X \in A]$ is all of $\mathcal{H}$ is called the spectrum of X. It is the complement of the set of points x such that $X - x\mathbf{1}$ has a bounded inverse. This includes any eigenvalues of X since if x is an eigenvalue corresponding to the eigenvector ψ, we have $(X - x\mathbf{1})\psi = \mathbf{0}$.

The so-called operator calculus we have described here extends to functions of several *commuting* operators. To avoid the nuisance caused by domain problems we define commutativity of two operators to mean that their projector-valued measures commute, i.e,

$$\Pi_{[X \in B]}\Pi_{[Y \in B']} \;=\; \Pi_{[Y \in B']}\Pi_{[X \in B]} \tag{3.14}$$

for all Borel sets B and B'. It follows that we do have $XY\psi = YX\psi$ for all ψ where both sides are defined, but this property alone does not imply commutativity. If two operators X and Y commute it can be easily seen that the product $\Pi_{[X \in B]}\Pi_{[Y \in B']}$ is also a projector, which we denote by $\Pi_{[X \in B, Y \in B']}$. These projectors form a projector-valued probability measure on the Borel rectangles $B \times B'$ of $\mathbb{R}^2$. It can be extended to a projector-valued measure on all the Borel sets of $\mathbb{R}^2$, which we denote by $A \mapsto \Pi_{[(X,Y) \in A]}$. For a Borel function f from $\mathbb{R}^2$ to $\mathbb{R}$ we can now define a new operator

$$Z \;=\; f(X, Y) \tag{3.15}$$

(which commutes with both X and Y) by any of the three routes:

$$\langle \psi | f(X,Y)\psi \rangle \;=\; \int\int f(x,y) \langle \psi | \Pi_{[X \in \mathrm{d}x, Y \in \mathrm{d}y]}\psi \rangle; \tag{3.16}$$

$$\Pi_{[Z \in B]} \;=\; \Pi_{[(X,Y) \in f^{-1}(B)]}; \tag{3.17}$$

$$f(X,Y) \;=\; \int\int f(x,y)\Pi_{[X \in \mathrm{d}x, Y \in \mathrm{d}y]}. \tag{3.18}$$

A theorem of von Neumann states that a collection of self-adjoint operators are pairwise commuting if and only if they are all functions of one self-adjoint operator.

If an operator has a discrete (countable) spectrum then the locations $x \in \mathbb{R}$ of the atoms of its projector-valued measure are its eigenvalues and the subspaces $[X = x]$ are the corresponding eigenspaces. An operator is bounded if and only if its spectrum is bounded.

Example 5 (Observable with continuous spectrum). Define $\mathcal{H}$ as the set of square integrable (measurable) complex functions on the real line; functions which are Lebesgue almost everywhere equal to one another being identified. We can represent a vector $\psi \in \mathcal{H}$ by some function $a \mapsto \psi(a)$, $a \in \mathbb{R}$, where we use a as argument in order to distinguish between possible values x of observables; there is not necessarily any connection between the two! The space $\mathcal{H}$ is separable. For instance, the usual Haar functions or trigonometric functions provide countable bases for the subspaces of functions zero outside of a bounded interval; these can be put together over the intervals $[j, j+1]$, integer j. Consider the operator X which is pointwise multiplication of the function ψ by the identity function $\iota : a \mapsto a$. The projector valued measure associated with X is determined by $\int x \|\Pi_{[X \in \mathrm{d}x]}\psi\|^2 = \langle\psi|X\psi\rangle = \int a|\psi(a)|^2 \mathrm{d}a$ for ψ such that $\int a^2|\psi(a)|^2 \mathrm{d}a < \infty$, indeed a dense subset of $\mathcal{H}$. Inspection of this equation suggests that $\Pi_{[X \in B]}$ is the operator which projects onto the functions zero outside of B. If this conjecture were true, then $\langle\psi|\Pi_{[X \in B]}\psi\rangle$ would equal $\int_B |\psi(a)|^2 \mathrm{d}a$. The measure $\langle\psi|\Pi_{[X \in \mathrm{d}x]}\psi\rangle$ would then be $|\psi(x)|^2 \mathrm{d}x$ making $\int x \|\Pi_{[X \in \mathrm{d}x]}\psi\|^2$ equal, as we hoped, to $\int a|\psi(a)|^2 \mathrm{d}a$. This equality (for all ψ) characterizes $\Pi_{[X \in B]}$, so our conjecture is correct.

The spectrum of X is all of $\mathbb{R}$; however X has no eigenvectors (and no eigenvalues) since if $X\psi = x\psi$ we would have $a\psi(a) = x\psi(a)$ for all a which would make $\psi(a)$ zero for almost all a, i.e., ψ would be the zero element of $\mathcal{H}$. The situation can in a sense be saved by using generalised functions (the Dirac delta functions then form a continuum of eigenfunctions) leading to so-called rigged Hilbert spaces, but these rather involved and delicate technical modifications do not make any difference to the results we want to use.

Since the subspace $[X \in B]$ is the space of functions zero outside of B, we see that $[X = x] = [X \in \{x\}]$ is actually the trivial subspace of functions zero almost everywhere; in other words, the zero function.

Replacing X by a bounded one-to-one function of itself, e.g., $\arctan(X)$, supplies an example of a bounded self-adjoint operator with continuous spectrum.

3.3 Observables and states.

Having fixed the notion of a (possible unbounded) self-adjoint operator and its projector-valued measure we are now able to carry over the theory of discrete quantum systems to the general continuous case. The state space of a continuous quantum system is a complex Hilbert space $\mathcal{H}$ with countable orthonormal basis, i.e., $\mathcal{H}$ is infinite-dimensional but separable.

We identify the physical notion of observable (or physical quantity) of the system with self-adjoint operator. We identify the state of the system with a non-negative self-adjoint operator ρ of trace 1, also called a *density matrix*. The trace of an operator X is by definition equal to $\sum\langle\phi_i|X\phi_i\rangle$ where (ϕ_i) is an orthonormal basis of $\mathcal{H}$; if this sum is absolutely convergent then it does not depend on which basis we take, and the operator is called a trace-class operator. Products of bounded self-adjoint operators and self-adjoint trace-class operators are again trace class. Inside a trace we may cyclically permute operators.

Self-adjoint trace-class operators (self-adjoint operators of finite trace) always have discrete spectrum and therefore any density matrix can be written as $\rho = \sum p_i \Pi_{[\phi_i]}$ where the ϕ_i form an orthonormal basis of $\mathcal{H}$ and the p_i are a probability distribution. Any projector onto a one-dimensional subspace is a density matrix and any (probabilistic) mixture of density matrices is again a density matrix. A state ρ which is a one-dimensional projector is called a pure state; all others are called mixed. A mixed state can always be represented as a discrete mixture of pure states. A density matrix ρ has a square root $\rho^{1/2}$ which is also non-negative. Hence for any non-negative self-adjoint operator X and a density matrix ρ we have $\mathrm{trace}(\rho X) = \mathrm{trace}(\rho^{1/2} X \rho^{1/2}) \geq 0$.

Consider a quantum system in state ρ and an observable (self-adjoint operator) X. Since $\Pi_{[X\in B]}$ is a nonnegative self-adjoint matrix we have that $\mathrm{trace}(\rho\Pi_{[X\in B]}) \geq 0$. Since it forms a projector-valued probability measure and since $\mathrm{trace}(\rho) = 1$ it follows that $B \mapsto \mathrm{trace}(\rho\Pi_{[X\in B]})$ is an ordinary probability measure on the real line. In fact if $\rho = \sum p_i \Pi_{[\phi_i]}$ then using the fact that $\Pi_{[\phi_i]} = \phi_i \phi_i^*$ we can calculate

$$\sum p_i \,\mathrm{trace}(\phi_i \phi_i^* \Pi_{[X\in B]}) = \sum p_i \|\Pi_{[X\in B]}\phi_i\|^2. \tag{3.19}$$

We continue therefore to make the interpretation that measurement of the observable X on a system in state ρ produces values x distributed according to the probability distribution $\Pr\{X_\rho \in B\} = \mathrm{trace}(\rho\Pi_{[X\in B]})$. These values lie in the *spectrum* of the corresponding operator. It is possible that X has a continuous spectrum; then X has no eigenvalues and the subspaces $[X = x]$ are the zero subspace for each value x. Projection of a pure state into that

subspace and renormalisation does not make sense. It seems reasonable to interpret this by saying that *exact* measurement of an observable with continuous spectrum is impossible without destroying the system. However for an observable with discrete spectrum (and hence eigenvalues and eigenspaces), $X = \sum x_i \Pi_{[X=x_i]}$, we may keep to our interpretation that when the value x is found, which occurs with probability $\mathrm{trace}(\rho\Pi_{[x=x]})$, the system is now in the state $\Pi_{[x=x]}\rho\Pi_{[x=x]}/\,\mathrm{trace}(\rho\Pi_{[x=x]})$.

Everything else we said about discrete quantum systems generalises immediately to continuous ones. In particular we emphasize that the laws of the unconscious quantum statistician, concerning the preservation of functional constraints between commuting observables in simultaneous measurements of those observables, remain valid. The tensor product construction of interacting systems built up from subsystems is still available. The notion of generalised observable or generalised measurement, replacing *projector-valued* probability measures by (self-adjoint) *operator-valued* probability measures carries over unchanged. The result characterizing all generalised measurements of a quantum system with an operator-valued measure remains true, and the Naimark extension theorem still allows any operator-valued measure to be represented as a quantum randomisation. Time evolution according to Schrödinger's equation takes the same form as before.

We include here a proof of one form of the law of the unconscious quantum statistician. We could not locate an explicit statement or proof of this result in the literature, except in either highly abstract or, alternatively, in rather restrictive forms. This is rather surprising since it is both practically and conceptually important, and it is often used without argument. We use the following notation: $(X_{\rho:X,Y,Z}, Y_{\rho:X,Y,Z}, Z_{\rho:X,Y,Z})$ denotes a triple of random variables with the *joint* distribution of a *simultaneous* measurement of three compatible observables X, Y and Z on a quantum system in state ρ; and similarly for other numbers of compatible observables. The theorem is concerned with properties of, and relationships between, such probability distributions; but it is easier to state in terms of random variables having these distributions. The probability spaces on which they are defined are irrelevant.

Theorem 1 (Preservation of functional constraints). *Suppose*

$$Z = f(X, Y) \tag{3.20}$$

where X and Y commute, then

$$Z_{\rho:X,Y,Z} = f(X_{\rho:X,Y,Z}, Y_{\rho:X,Y,Z}) \qquad \textit{with probability } 1. \tag{3.21}$$

As a corollary we have

$$Z_\rho \;=\; f(X_{\rho:X,Y}, Y_{\rho:X,Y}) \qquad \textit{in distribution.} \tag{3.22}$$

Proof. The steps in the proof are as follows, some in fact already introduced above. We are given that the projector-valued measures of X and Y commute, so we can define a projector-valued measure on the Borel rectangles of $\mathbb{R}^2$ by $B \times B' \mapsto \Pi_{[X\in B]}\Pi_{[Y\in B']}$. We can extend this to a unique projector-valued measure on all of the Borel sets of $\mathbb{R}^2$, which we denote by $C \mapsto \Pi_{[(X,Y)\in C]}$. Next we define a new projector-valued measure on $\mathbb{R}$ by $B \mapsto \Pi_{[(X,Y)\in f^{-1}(B)]}$. This projector-valued measure corresponds to a unique observable Z in the standard way: $[Z \in B] = [(X,Y) \in f^{-1}(B)]$. The new observable commutes with the two given ones and the joint distribution of measurements of all three is determined by its values on the Borel rectangles of $\mathbb{R}^3$ which are

$$\begin{aligned}
\Pr\{X_{\rho:X,Y,Z} \;&\in\; B, Y_{\rho:X,Y,Z} \in B', Z_{\rho:X,Y,Z} \in B''\} \\
&= \; \mathrm{trace}(\rho\Pi_{[X\in B]}\Pi_{[Y\in B']}\Pi_{[Z\in B'']}) \\
&= \; \mathrm{trace}(\rho\Pi_{[(X,Y)\in B\times B']}\Pi_{[(X,Y)\in f^{-1}(B'')]}) \\
&= \; \mathrm{trace}(\rho\Pi_{[(X,Y)\in (B\times B')\cap f^{-1}(B'')]}) \\
&= \; \Pr\{(X_{\rho:X,Y}, Y_{\rho:X,Y}) \in (B\times B') \cap f^{-1}(B'')\} \\
&= \; \Pr\{X_{\rho:X,Y} \in B, Y_{\rho:X,Y} \in B', f(X_{\rho:X,Y}, Y_{\rho:X,Y}) \in B''\}.
\end{aligned} \tag{3.23}$$

This is the same as the distribution of $(X_{\rho:X,Y}, Y_{\rho:X,Y}, f(X_{\rho:X,Y}, Y_{\rho:X,Y}))$ and is therefore concentrated on $\{z = f(x,y)\}$ in $\mathbb{R}^3$. □

Let us summarize the main points of this chapter.

3.3.1 Summary.

1. Observables (physical quantities) are represented by self-adjoint operators X. Self-adjoint operators are in a one-to-one correspondence with projector-valued measures. The correspondence can be expressed as $X = \int x\Pi_{[X\in \mathrm{d}x]}$. This representation yields the distribution of values found when X is measured on a system in state ρ, as we will describe in the following items.

2. States are represented by self-adjoint, non-negative, unit trace operators ρ. Such an operator is called a density matrix. It can be written (generally in many ways) as $\rho = \int \Pi_{[\psi]} P(\mathrm{d}\psi)$ where $P(\mathrm{d}\psi)$ is a probability distribution over normalized state vectors ψ. If this mixing distribution is degenerate the state is called pure, otherwise it is called

mixed. A mixed state can always be written as a discrete mixture of one-dimensional orthogonal projectors.

3. With X_ρ denoting a random variable having the distribution found when X is measured on a system in state ρ, we have $\Pr\{X_\rho \in B\} = \operatorname{trace}(\rho \Pi_{[X \in B]})$. Consequently $\mathrm{E}(X_\rho) = \operatorname{trace}(\rho X)$ if the expectation—or the trace—exists.

4. If X has a discrete spectrum, i.e., $X = \sum x \Pi_{[X=x]}$, then if on measurement of X the value x is found, the system is now in the state $\Pi_{[X=x]} \rho \Pi_{[X=x]} / \operatorname{trace}(\rho \Pi_{[X=x]})$.

5. For a Borel function f, by definition $f(X) = \int f(x) \Pi_{[X \in \mathrm{d}x]}$; it follows that $(f(X))_\rho = f(X_\rho)$ in distribution (the simplest form of the law of the unconscious quantum statistician).

Example 6 (Classical as special case of quantum). Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Suppose $\mathcal{F}$ is countably generated. Now define the separable Hilbert space $\mathcal{H}$ to be the space of absolute-square-integrable *complex-valued* random variables. Let Θ be any real random variable and let ρ be the pure state corresponding to $e^{i\Theta}$. Each *real* random variable X on the original probability space corresponds to an observable, or self-adjoint operator X, defined as follows: it acts on $\psi \in \mathcal{H}$ by pointwise multiplication of the values $(\psi(\omega) : \omega \in \Omega)$ by the values $(X(\omega) : \omega \in \Omega)$ to produce the complex random variable $\omega \mapsto X(\omega)\psi(\omega)$. The projector-valued measure of the observable X corresponds to the subspaces $[X \in B]$ of square-integrable complex random variables which are zero outside of the event $\{X \in B\}$. The observables X defined in this way all commute and the joint distribution of measurements of them is exactly their distribution as random variables on $(\Omega, \mathcal{F}, \mathrm{P})$. Many other observables also exist of course.

Chapter 4

Chapters to come ...

4.1 Parametric estimation

We concentrate on the quantum Cramér-Rao bound and its application to estimation of the spin half state, given n identical copies of the system; see Barndorff-Nielsen and Gill (1998, 2000), Gill (2000), Gill and Massar (2000).

4.2 Exponential families and transformation models

Various generalisations of the notion of exponential family. Transformation models. Relations and properties.

4.3 Quantum tomography

Nonparametric estimation of the state of an infinite dimensional system. The quantum simple harmonic oscillator. This material continues from the chaper on infinite dimensional quantum systems.

4.4 Quantum stochastic processes

Main theme: evolutions of a density matrix modelled as continuous measurement.

4.5 Future quantum technology

A survey of the basic ideas in quantum computing, communication, information, teleportation and cryptography.

Chapter 5

Hidden variables

This chapter surveys some celebrated results showing that any deterministic explanation for the randomness apparent in quantum measurements will of necessity possess some highly unattractive features. A theory purporting to underlie quantum mechanics is called a 'hidden variables theory'. The aim of such a theory is to express the phenomena *described* so effectively by quantum theory as just the surface expression of a more fundamental level, to which, hopefully, more classical physical laws apply. The mathematical results show on structural grounds that such programmes will fail. Hidden variables models can be constructed (and work on them continues today) but the models exhibit of necessity features even more surprising than those of quantum mechanics itself.

Research in this direction centres around the famous Bell (1967) inequalities: these are inequalities which should hold for any model satisfying some apparently harmless physical constraints, but which are violated by quantum mechanical predictions. The violation has been confirmed experimentally, in particular, in the famous Aspect experiment, Aspect et al. (...). Earlier central results in this area are the fundamental theorems of Gleason (1957) and Kochen and Specker (1967). In recent years various authors have derived further Bell-type paradoxes, in a sense 'almost-sure' versions of Bell's result, and we will also discuss one of these; the so-called Mermin (...) array. Bell's example, and even more clearly these later examples, show that already the surface phenomena of quantum mechanics violate cherished physical principles, independently of any attempt to 'explain' quantum phenomena.

5.1 Kochen-Specker vs. noncontextual h.v.'s

Could the randomness in the outcome of measurement of any observable on a quantum system in a given state ρ simply be the expression of statistical variability in further, 'hidden', state variables? In other words, when we say that the system is in state ρ, in fact the real state of the system is not completely fixed. There are other variables which we are not able to control or keep constant, variation in which completely explains the variability of measurement results.

We first discuss the most simplistic type of hidden variables model. Keeping a particular quantum state ρ fixed in the following discussion, let us denote by ω a possible configuration (set of values) of all necessary hidden variables. Let Ω denote the collection of all possible ω. Measurement of the observable X on the system in hidden state ω results in a value which is a deterministic function of X and ω, which we will denote by $X_\rho(\omega)$. This notation makes explicit a key assumption that this value does not depend on which other (compatible) observables are being measured at the same time as X. In other words, the hidden variables truly reside in the quantum system under study, not in the apparatus which is employed to measure it. Such a hidden variables model is called *non-contextual*. A non-contextual hidden variables model can be interpreted to say that the result of the measurent of the observable X already exists, independently of the measurement, as the value $x = X_\rho(\omega)$. Measurement simply reveals this pre-existing value.

Indeed such a model seems very simplistic. Measurement devices are built up of very, very many quantum subsystems, and if there is variability in a single fundamental particle which we cannot control, then surely even more so in macroscopic measuring devices. However nature itself gives a clear signal that we should entertain such non-contextual models: the probability distribution of a measurement of X when it is measured jointly with a compatible variable Y, is exactly the same as when it is measured jointly with a different compatible variable Z, even though Y and Z are incompatible, so that completely different experimental set-ups are required for the two experiments. In other words, *statistically* the context in which X is measured makes *absolutely no difference* to the results, so why should this also not be true at the level of individual outcomes.

Recall that we denote by $X_\rho(\omega)$ the result of measuring X on a system in the hidden state ω. From quantum mechanics we can write down the joint probability distribution of measurements of any collection of *compatible* observables X, Y, ... According to our hidden variables model this joint distribution is the same as that inherited by $(X_\rho(\cdot), Y_\rho(\cdot), \dots)$ from the variability in the hidden argument ω.

A modern-day probabilist or statistician will formulate this mathematically as follows: there exists a probability measure P on the space of hidden states Ω, on which a sigma-algebra $\mathcal{F}$ is defined making the functions X_ρ measurable. Under this probability measure, the joint distribution of any set $(X_\rho, Y_\rho, \dots)$ corresponding to compatible observables $(X, Y, \dots)$ is the one predicted by quantum mechanics,

$$\mathrm{P}(X_\rho \in B, Y_\rho \in C, \dots) = \mathrm{trace}(\rho \Pi_{[X \in B]} \Pi_{[Y \in C]} \dots). \tag{5.1}$$

We should be careful however not to make our argument heavily dependent on technical issues such as measurability. We will show that the non-contextual hidden variables programme is doomed to failure already if we only consider a finite number of observables, each taking with probability one a finite number of different values. This implies that our sigma-algebra $\mathcal{F}$ of measurable sets can be taken to be finite, and the question is whether the values of P implied by (5.1) for various events in $\mathcal{F}$ (involving only compatible observables) can be extended consistently (additively, nonnegatively) to the rest of this algebra, in other words, to events involving arbitrary combinations of our finite collection of observables.

As such we can connect the hidden variables programme to the ordinary probabilistic notion of coupling. We are given a collection of probability spaces, each one modelling an experiment in which a set of compatible observables are measured. Can we construct a coupling of these spaces, i.e., a single probability space on which there is a single random variable corresponding to each observable, and such that the marginal distributions of sets of variables corresponding to compatible observables reproduces the right hand side of (5.1)?

The Kochen-Specker theorem which we are going to state, and in the next section prove, implies that from dimension 3 onwards, one can exhibit a finite collection of discrete observables which cannot be coupled in this way. This kind of result is called a *no-go theorem*, prohibiting a certain type of hidden variables model—in this case all non-contextual models. First we make some preparatory arguments, interpreting the problem in terms of 0–1 valued observables (projectors), projecting onto one-dimensional subspaces of a 3-dimensional subspaces of $\mathcal{H}$.

Now if a collection of compatible observables satisfying a functional relation $X = f(Y, Z, \dots)$ are measured simultaneously, we know that with probability one the outcomes will satisfy $X_\rho(\omega) = f(Y_\rho(\omega), Z_\rho(\omega), \dots)$. The outcomes are also contained in the spectrum of the observables with probability one. Now suppose $\dim(\mathcal{H}) \geq 3$. Then there exists a subspace A of $\mathcal{H}$ of dimension 3 such that $\mathrm{trace}(\rho \Pi_A) > 0$.

A projector is a 0–1 valued observable. The assumption $\mathrm{trace}(\rho\Pi_A) > 0$ means that a measurement of Π_A yields a value 1 with positive probability. If ψ, ψ' and ψ'' are orthogonal and span A, then

$$\Pi_A = \Pi_{[\psi]} + \Pi_{[\psi']} + \Pi_{[\psi'']} \tag{5.2}$$

and the four projectors commute. So a simultaneous measurement of the set of four yields with positive probability a 1 on the left hand side and two 0's and a 1 on the right hand side. Restrict attention to a finite collection of non-zero vectors $\psi_1, \psi_2, \ldots$ in A. We have that, for each ψ in the list, with probability 1,

$$(\Pi_{[\psi]})_\rho)(\omega) = 0 \text{ or } 1; \tag{5.3}$$

for each orthogonal triple, with probability 1,

$$(\Pi_A)_\rho(\omega) = (\Pi_{[\psi]})_\rho(\omega) + (\Pi_{[\psi']})_\rho(\omega) + (\Pi_{[\psi'']})_\rho(\omega); \tag{5.4}$$

and finally, with positive probability,

$$(\Pi_A)_\rho(\omega) = 1. \tag{5.5}$$

Since exceptions to (5.3) and (5.4) only hold with zero probability, and there are only a finite number of vectors ψ and orthogonal triples ψ, ψ' and ψ'', there must exist a point ω such that (5.3), (5.4) and (5.5) *all* hold for all vectors and triples under consideration.

In geometric language this says that any finite set of non-zero vectors in a three-dimensional subspace of $\mathcal{H}$ can be assigned the colours red (for 1) and green (for 0) subject to the constraint that in any orthogonal triple, there is one and only one assignment red.

The Kochen-Specker theorem states that there exist finite sets of vectors, for which such a colouring is not possible. Since on the other hand existence of a non-contextual hidden variables model implies the existence of a colouring for any finite set of vectors, we have as a corollary of the theorem that a non-contextual hidden variables model is impossible in any dimension from 3 onwards (and whatever the state ρ). For dimension 2, and for arbitrary *contextual* models, see the exercises at the end of this section!

Kochen and Specker proved their theorem by an explicit construction of 117 vectors in $\mathbb{C}^3$, for which a colouring satisfying the just mentioned constraints is impossible. Actually a similar construction can be extracted from the key paper of Bell(1966; ...). In both cases, the construction only needs $\mathbb{R}^3$, i.e., one only needs vectors which are *real* linear combinations of a given basis. Later authors have tried to 'improve' the theorem by reducing the

number of vectors; the current world record stands at 31 (an example due to Conway and Kochen, reported by Peres (1993)). Checking that the colouring is impossible is now a fairly involved combinatorial exercise. In the next section we will give a recent and geometrically inspired construction due to Gill and Keane (1996), which is similarly extravagant in terms of the number of vectors involved as the original Kochen-Specker or Bell arguments, but has the advantage that it can be 'seen' rather than relying on combinatorial or algebraic computations. Moreover its proof is built on a key lemma for the proof of the related Gleason theorem, which we will also discuss.

The reader may however skip that section and proceed with Section 14.3 on the Bell inequalities. These inequalities also yield a no-go result for a less restricted class of hidden-variables models, and therefore also contain an example of Kochen-Specker type. However dimension four, at least, and a special choice of state ρ, are needed to build this example. As such it exemplifies the Kochen-Specker theorem, but does not prove it.

We finish this section with two more technical remarks and an exercise. The Kochen-Specker theorem can be interpreted to be say that 'joint distributions of incompatible observables do not exist. In fact theorems with apparently such a content go back to the early days of quantum mechanics. For instance, there is a theorem stating that if X and Y are two incompatible obervables, then there is no joint distribution of a pair of random variables X_ρ, Y_ρ such that the marginal law of $\alpha X_\rho + \beta Y_\rho$ is equal to the distribution of measurements of the observable $\alpha X + \beta Y$ for all real α, β. However this result has no bearing on the issue of hidden variables, since there is no a priori reason why the observable $\alpha X + \beta Y$ should have any connection whatsoever with the conjectured random variable $\alpha X_\rho + \beta Y_\rho$.

Pitowsky (...) has claimed that the whole problem is solved by allowing the X_ρ not to be measurable, but in our opinion this is not the correct interpretation of his results, which merely show that there is no problem if one leaves probabilities undefined which otherwise would be forced to be negative. He manages to disguise this 'solution' by a most elaborate and technical construction involving laws of large numbers for non-measurable random variables, but the heart of his argument is that if two observables are not compatible, then probabilities involving values of both *should* be undefined. This completely avoids the whole issue.

Problem 13 (Hidden variables for dimension 2). Construct a hidden variables model for a given state on two-dimensional Hilbert space.

Problem 14 (Product construction). A given collection of probability spaces can always be coupled, e.g., the product construction yields a single

probability space on which all of the random variables on each of the component spaces exist, and therefore have joint distributions. Show that this construction supplies a trivial (though physically unattractive) *contextual* hidden variables theory in which $X_{\rho:X,Y,\ldots}(\omega)$ denotes the result of measuring the observable X, in the company of compatible observables Y, Z, ..., when the quantum system is in the hidden state ω.

5.2 Gleason's theorem

Fixing a state ρ on a quantum system, consider the mapping p from closed subspaces of $\mathcal{H}$ to the interval $[0,1]$ defined by

$$p(A) = \mathrm{trace}(\rho\Pi_A). \tag{5.6}$$

This mapping associates a probability with each subspace, namely the probability that measurement of the binary (or Bernoulli) observable Π_A would yield the value 1. Thinking of the subspaces A as 'properties of the quantum system', we have defined a probability measure on the spaces of properties in the sense that $p(A)$ is the probability that the system 'would be found to possess property A', if that property were measured. Now the term 'probability measure' is quite well justified because it is easy to check that p satisfies the usual axioms of a probability measure if we draw the following analogies: orthogonality of subspaces and disjointness of events; the orthosum of orthogonal subspaces and the union of disjoint events; the zero subspace and the empty event; the whole space and the certain event. Thus (with some redundancy)

$$0 \leq p(A) \leq 1; \tag{5.7}$$

for any countable collection of orthogonal A_i

$$p(\oplus A_i) = \sum p(A_i); \tag{5.8}$$

and

$$p(\{0\}) = 0, \quad p(\mathcal{H}) = 1. \tag{5.9}$$

Gleason's (1956) theorem states that *any probability measure p on the lattice of subspaces of a separable Hilbert space has to be of the form* $p(A) = \mathrm{trace}(\rho\Pi_A)$ *for some density matrix* ρ. This result is fundamental in a long-lasting research programme to derive the mathematical structure of quantum mechanics from a more appealing, more qualitative set of axioms. If one could

argue from some general principles that 'properties' of a quantum system must of necessity have the same algebraic structure as subspaces of a Hilbert space, and their 'probabilities' must satisfy the above axioms, then one has derived the whole Hilbert space set-up with operators and the trace operation as a necessary consequence of the initial axioms.

Gleason's theorem (already conjectured by von Neumann) has a long history. The first proofs of it (Piron ...) were not entirely rigorous and Gleason's own proof was very long and complicated; it used difficult methods from harmonic analysis. An 'elementary' proof was first given by Cooke, Keane and Moran (1985). This proof needs fairly elementary analysis only, together with a a geometric lemma due to Piron. We will use exactly this lemma in our proof of the Kochen-Specker theorem. After that we will prove that theorem, and then comment on some other links between the Kochen-Specker and the Gleason theorems.

5.3 A geometric lemma

Consider the one-dimensional subspaces corresponding to non-zero, real, linear combinations of three orthogonal vectors in $\mathbb{C}^k$, $k \geq 3$. These subspaces may be represented by points on (the surface of) the Northern hemisphere of the globe. The original triple is represented by North pole together with two points on the equator whose longitudes differ by 90°.

Now fix a point ψ in the Northern hemisphere, not at the North pole nor on the equator. Consider the great circle through this point which crosses the equator at the two points differing in longitude by $\pm 90°$ from ψ. Choose one of these equatorial points and call it ψ^{E}. Call the point on the Northern hemisphere orthogonal to the great circle $\psi^{\perp}$. Its longitude is that of ψ plus 180° and its latitude is 90° minus that of ψ. The triple ψ, ψ^{E}, $\psi^{\perp}$ are orthogonal.

The great circle we just defined has ψ as its most Northerly point. We call it *the great circle descent from* ψ.

Starting from a point $\psi = \psi_0$ go down its descent circle some way to a new point ψ_1. Now consider the new great circle descent from ψ_1. Go down some way to a new point ψ_2, and so on. After n steps arrive at ψ_n. Obviously ψ_n is more Southerly than ψ_0. Cooke, Keane and Moran's geometric lemma states that one can reach *any* more Southerly point than ψ_0 by a finite sequence of great circle descents. For instance, one can fly from Amsterdam to Tokyo by a finite sequence of great circle descents.

The lemma is proved by projecting the Northern hemisphere *from* the centre of the earth *onto* the horizontal plane tangent to the earth at the North

pole. Circles of constant latitude project onto concentric circles, a great circle descent projects onto a straight line tangent to the circle of constant latitude at its summit.

PICTURE

5.4 Proof of the Kochen-Specker theorem

The theorem is proved by exhibiting a finite collection of unit vectors in $\mathbb{R}^3$ such that it is impossible to colour each vector either red or green subject to the following constraints: 1), within any orthogonal triple, exactly one vector is red and the other two are green; 2), if one vector lies in the subspace generated by two orthogonal vectors and those two are both coloured green, then the third is coloured green as well. The second constraint follows from the first by constructing a vector orthogonal to the first two. It must be coloured red. Call these three vectors ψ_1, ψ_2 and ψ_3, where ψ_1 is red and ψ_2 and ψ_3 are green. We are interested in a vector ϕ in the plane generated by ψ_2 and ψ_3. It is automatically orthogonal to ψ_1 (red) and there exists another vector, say η, in the plane generated by ψ_2 and ψ_3, and orthogonal to ϕ. Now ψ_1, ϕ and η form an orthogonal triple, while ψ_1 is already coloured red. The other two, in particular ϕ, therefore have to be green.

Now we start the construction. Fix an an orthogonal triple. Colour one point red and the other two green. Let the red point be the North pole and the other two green points be on the equator. Any further points selected on the equator get coloured green by rule 2). Take a point ψ at latitute 60° above the equator. Together with $\psi^\perp$ and ψ^{E} we have a new orthogonal triple. Since ψ^{E} gets coloured green, if ψ is coloured green then $\psi^\perp$ is coloured red. Note that $\psi^\perp$ lies at 30° above the equator, more Southerly than ψ.

Suppose ψ is coloured green. Since any point on its great circle descent is a linear combination of ψ and ψ^{E}, it is also coloured green. Repeating this argument, any point which can be reached by a finite number of great circle descents from ψ is also coloured green. But this applies to $\psi^\perp$, a contradiction.

Therefore ψ is coloured red just like the North pole. So we have shown that any point within 30° of a red point is also coloured red. Now go in three steps of 30° from the North pole down to the equator, then in three steps of 30° along the equator, then in three steps of 30° back up to the North pole. One of the three 'corners' of this circuit has to be coloured red, hence they all are, a contradiction. □

It has been claimed that Gleason's theorem itself supplies a very direct and short proof of the no-go theorem for noncontextual hidden variables

models. The putative proof is as follows. Fix a point ω and consider all evaluations $e(A) = (\Pi_A)_\rho(\omega)$. Almost surely, each evaluation takes the values zero and one. Almost surely, $e(A)$ is additive on disjoint subspaces, zero on the zero subspace and equal to one on the whole space. If the point ω is not in any of the exceptional (zero probability) events, then e is a probability measure on the lattice of subspaces of $\mathcal{H}$ and hence of the form $\mathrm{trace}(\rho\Pi_A)$ for some density matrix A. However such a probability measure takes values strictly between zero and one on some subspaces A, while e only takes these values.

Unfortunately this argument requires us to exclude uncountably many null sets, and these could exhaust Ω. It does not seem easy to repair this.

5.5 Bell vs. contextual local hidden variables

The results above show that hidden-variables models have to be contextual. Hidden variables must reside in the measurement apparatus as well as in the quantum system under investigation. The result of measuring one particular observable depends not only on the quantum system but also on the measurement apparatus, and in particular depends on what other observables are being measured at the same time.

Now if we model two separate particles as a single quantum system using the tensor product construction, we can also consider an observable on one of the subsystems as an observable on the joint system. Two observables, one on each subsystem, are compatible. Now in some physical examples, the two particles, at the time of measurement, might be widely separated in space, so that measurement on one particle can be done 'independently' of measurement of the other. In particular the *result* of measuring one of the two particles should not depend on *which* observable is measured on the other.

Let X and X' represent two incompatible observables on the first subsystem, and Y and Y' represent two incompatible observables on the second. However X is compatible with Y and with Y', and X' is also compatible with both Y and Y'. There are now four different experiments which might be carried out: measure X and Y, or measure X and Y', or measure X' and Y, or measure X' and Y'. A contextual hidden variables model will specify four pairs of random variables, let us write for example $X_{\rho:X,Y}(\omega)$ for the result of measuring X on the first subsystem when simultaneously Y is measured on the second subsystem, where ω stands for all hidden variables needed to specify the outcome of any of the possible experiments. There is no problem in finding a probability space on which all these four pairs are

defined simultaneously with the right probability distributions.

There may however be a problem if want to impose the natural restriction stated above. That the result of measuring one subsystem should not depend on which measurement is made on the other, translates into $X_{\rho:X,Y}(\omega) = X_{\rho:X,Y'}(\omega)$ and three other identities. There are now just four random variables, which we abbreviate to X_ρ, X'_ρ, Y_ρ and $Y'\rho$. We ask the question: can we construct a single probability space with just these four random variables such that each of the four pairs (X_ρ or X'_ρ together with Y_ρ or Y'_ρ) have the same bivariate distribution as that of joint measurements of the corresponding pair of compatible observables (X or X' together with Y or Y')?

Let us specialise further and suppose that the observables under consideration are all binary. Let us suppose that the coupling is possible. This puts some constraints on the bivariate distributions. For instance, the marginal distribution of one variable can be computed from two different pairs, and these must give the *same* answer; for instance, X_ρ occurs in the pair (X_ρ, Y_ρ) and in (X_ρ, Y'_ρ). This constraint will be satisfied, as we know already. We will derive a further *inequality* relating the distributions of the four pairs. Then we will turn to a specific example, and show that the inequality is violated.

Fix a point ω and consider the four values $x = X_\rho(\omega)$, $x' = X'_\rho(\omega)$, $y = Y_\rho(\omega)$ and $y' = Y'_\rho(\omega)$. These four values are all 0 or 1. Moreover it is easy to verify that

$$x \neq y' \;\&\; y' \neq x' \;\&\; x' \neq y \quad \Longrightarrow \quad y \neq x; \tag{5.10}$$

simply fill in (without loss of generality) $x = 0$ and calculate the value of y. Conversely we must have

$$x = y \quad \Longrightarrow \quad x = y' \text{ or } y' = x' \text{ or } x' = y. \tag{5.11}$$

Therefore we obtain Bell's inequality

$$\mathrm{P}(X_\rho = Y_\rho) \leq \mathrm{P}(X_\rho = Y'_\rho) + \mathrm{P}(X'_\rho = Y'_\rho) + \mathrm{P}(X'_\rho = Y_\rho). \tag{5.12}$$

One can write down three other inequalities of this type and it turns out (Fine, ...) that a joint distribution of all four binary variables exists if and only if the univariate marginals match (four equality constraints) and these four inequalities are satisfied.

Let us now consider a specific quantum model. Polarization measurements on a coupled pair of photons can be modelled in the four dimensional space $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$. Denote the unit vector $(\cos\theta, \sin\theta)^\top$ in $\mathbb{C}^2$ by $|\updownarrow_\theta\rangle$. Write $\Pi_\theta = |\updownarrow_\theta\rangle\langle\updownarrow_\theta|$. Define furthermore two basis vectors $|\updownarrow\rangle = |0\rangle$ and

$|\leftrightarrow\rangle = |\pi/2\rangle$. Recall that these vectors describe the states of a single photon polarized vertically and polarized horizontally respectively. We write product states for instance as $|\updownarrow\leftrightarrow\rangle = |\updownarrow\rangle \otimes |\leftrightarrow\rangle$. Note that

$$\langle\updownarrow_\theta|\updownarrow_\phi\rangle = \cos\theta\cos\phi + \sin\theta\sin\phi = \cos(\theta-\phi). \tag{5.13}$$

Consider a pair of photons in the pure state corresponding to the unit vector, called the singleton or Bell state

$$|\mathrm{S}\rangle = \tfrac{1}{2}(|\updownarrow\leftrightarrow\rangle - |\leftrightarrow\updownarrow\rangle). \tag{5.14}$$

The corresponding density matrix is

$$\begin{aligned}\rho &= |\mathrm{S}\rangle\langle\mathrm{S}| \\ &= \tfrac{1}{2}\big((|\updownarrow\rangle\langle\updownarrow|)\otimes(|\leftrightarrow\rangle\langle\leftrightarrow|) \\ &\quad -(|\updownarrow\rangle\langle\leftrightarrow|)\otimes(|\leftrightarrow\rangle\langle\updownarrow|) \\ &\quad -(|\leftrightarrow\rangle\langle\updownarrow|)\otimes(|\updownarrow\rangle\langle\leftrightarrow|) \\ &\quad +(|\leftrightarrow\rangle\langle\leftrightarrow|)\otimes(|\updownarrow\rangle\langle\updownarrow|)\big)\end{aligned} \tag{5.15}$$

where we used the bilinearity and the rule $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$. The projector for the event 'photon 1 passes through a polarization filter aligned at angle θ to the vertical is $\Pi_\theta \otimes \mathbf{1}$, and similarly, the projector for 'photon 2 passes through a polarization filter aligned at angle ϕ to the vertical is $\mathbf{1}\otimes\Pi_\phi$'. These two projectors commute and their product is $\Pi_\theta\otimes\Pi_\phi$, projector for the event 'photons 1 and 2 simultaneously pass through filters at angles θ and ϕ.

The joint distribution of a measurement of these two is determined by the marginal probabilities $\mathrm{trace}(\rho\Pi_\theta\otimes\mathbf{1})$ and $\mathrm{trace}(\rho\mathbf{1}\otimes\Pi_\phi)$, and the joint probability $\mathrm{trace}(\rho\Pi_\theta\otimes\Pi_\phi)$. Now,

$$\begin{aligned}\mathrm{trace}(\rho\Pi_\theta\otimes\mathbf{1}) = \tfrac{1}{2}\Big\{ &\ \mathrm{trace}(|\updownarrow\rangle\langle\updownarrow|\otimes|\leftrightarrow\rangle\langle\leftrightarrow|\ \Pi_\theta\otimes\mathbf{1}) \\ &- \mathrm{trace}(|\updownarrow\rangle\langle\leftrightarrow|\otimes|\leftrightarrow\rangle\langle\updownarrow|\ \Pi_\theta\otimes\mathbf{1}) \\ &- \mathrm{trace}(|\leftrightarrow\rangle\langle\updownarrow|\otimes|\updownarrow\rangle\langle\leftrightarrow|\ \Pi_\theta\otimes\mathbf{1}) \\ &+ \mathrm{trace}(|\leftrightarrow\rangle\langle\leftrightarrow|\otimes|\updownarrow\rangle\langle\updownarrow|\ \Pi_\theta\otimes\mathbf{1})\Big\}.\end{aligned} \tag{5.16}$$

Using the rule

$$\mathrm{trace}(A\otimes B\,C\otimes D) = \mathrm{trace}(AC\otimes BD) = \mathrm{trace}(AC)\,\mathrm{trace}(BD), \tag{5.17}$$

together with $\mathrm{trace}(|i\rangle\langle j|\,\mathbf{1}) = \mathrm{trace}(|i\rangle\langle j|) = \delta_{ij}$, two of the four terms here disappear. We also have $\mathrm{trace}(|\updownarrow\rangle\langle\updownarrow|\,\Pi_\theta) = \mathrm{trace}\,\Pi_0\Pi_\theta = \cos^2(\theta)$, and $\mathrm{trace}(|\leftrightarrow\rangle\langle\leftrightarrow|\,\Pi_\theta) = \mathrm{trace}\,\Pi_{\pi/2}\Pi_\theta = \sin^2(\theta)$. Thus the marginal probability we are computing reduces to $\frac{1}{2}(\cos^2(\theta) + \sin^2(\theta)) = \frac{1}{2}$. Similarly the other marginal probability is also $\frac{1}{2}$. Finally,

$$\begin{aligned}
\mathrm{trace}(\rho\Pi_\theta \otimes \Pi_\phi) = \tfrac{1}{2}\Big\{ & \mathrm{trace}(|\updownarrow\rangle\langle\updownarrow|\,|\updownarrow_\theta\rangle\langle\updownarrow_\theta|)\,\mathrm{trace}(|\leftrightarrow\rangle\langle\leftrightarrow|\,|\updownarrow_\phi\rangle\langle\updownarrow_\phi|) \\
& - \mathrm{trace}(|\updownarrow\rangle\langle\leftrightarrow|\,|\updownarrow_\theta\rangle\langle\updownarrow_\theta|)\,\mathrm{trace}(|\leftrightarrow\rangle\langle\updownarrow|\,|\updownarrow_\phi\rangle\langle\updownarrow_\phi|) \\
& - \mathrm{trace}(|\leftrightarrow\rangle\langle\updownarrow|\,|\updownarrow_\theta\rangle\langle\updownarrow_\theta|)\,\mathrm{trace}(|\updownarrow\rangle\langle\leftrightarrow|\,|\updownarrow_\phi\rangle\langle\updownarrow_\phi|)\Big) \\
& + \mathrm{trace}(|\leftrightarrow\rangle\langle\leftrightarrow|\,\psi_\theta\psi_\theta)\,\mathrm{trace}(|\updownarrow\rangle\langle\updownarrow|\,\psi_\phi\psi_\phi^*)\Big\} \\
= \tfrac{1}{2}\Big\{ & \cos\theta\cos\theta\sin\phi\sin\phi \\
& - \sin\theta\cos\theta\cos\phi\sin\phi \\
& - \cos\theta\sin\theta\sin\phi\cos\phi \\
& + \sin\theta\sin\theta\cos\phi\cos\phi\Big\} \\
= \tfrac{1}{2}(& \cos\theta\sin\phi - \sin\theta\cos\phi)^2 \\
= \tfrac{1}{2}\sin^2(\theta-\phi). &
\end{aligned} \tag{5.18}$$

The probabilities that each photon passes, or each photon does not pass, the two filters are therefore both $\frac{1}{2}\sin^2(\theta-\phi)$; the probabilities that one does and the other does not are both $\frac{1}{2}\cos^2(\theta-\phi)$. The probability that both photons do the same is therefore $\sin^2(\theta-\phi)$. If the two filters are aligned the same, exactly one of the photons will pass its filter; each choice with probability half. If the two filters are aligned at right angles then both the photons will pass or neither will pass, again, each with probability half. Marginally, each photon has probability half to pass any polarization filter.

Consider now a pair of filter settings for each of the two filters. We set $\theta = 0$, $\theta' = \pi/3$; $\phi = \pi/2$, $\phi' = \pi/6$. The angles θ, ϕ', θ', ϕ in sequence therefore increase in three steps of 30° from 0° to 90°. Define $X = \Pi_\theta \otimes \mathbf{1}$, $X' = \Pi_{\theta'} \otimes \mathbf{1}$, $Y = \mathbf{1} \otimes \Pi_\phi$, and $Y' = \mathbf{1} \otimes \Pi_{\phi'}$. Since $\sin^2(30°) = \frac{1}{4}$ while $\sin^2(90°) = 1$, we have $\mathrm{P}(X_\rho = Y_\rho) = 1$ while $\mathrm{P}(X_\rho = Y'_\rho) = \frac{1}{4}$, $\mathrm{P}(Y'_\rho = X'_\rho) = \frac{1}{4}$, and $\mathrm{P}(X'_\rho = Y_\rho) = \frac{1}{4}$. This violates Bell's inequality since $1 > \frac{1}{4} + \frac{1}{4} + \frac{1}{4}$.

A corresponding experiment was carried out by Alain Aspect and his coworkers in Orsay, Paris, 1992, and has since been replicated in many other places and in many forms. Apect considered ...

5.6 The Mermin array and other constructions

Our derivation of the Bell inequality made clear that it is not probability theory which leads to paradoxes with quantum theory, it is fairly innocent looking deterministic reasoning. The starting point was that 'the result of measuring X on the first particle does not depend on whether Y or Y' is measured on the second'. This point is made especially clear in Maudlin (1995), where the author asks you to consider the following experiment. You are in a room with a friend and you are about to each leave the room by a separate door. Outside the door you are each going to be asked a question and you may only give one of two answers, 'pass' or 'don't pass'. The question will actually specify an angle; one of you is going to be asked either 0° or 60°, the other is going to be asked 30° or 90°. You want to agree on a strategy such that, whenever the two questions at the two doors differ by 90°, you each give the same answer; but in any of the other three cases, you want your answers only to be the same in 25% of the time. You are allowed to randomise your answers using dice, cards or whatever; what you may not do is communicate in any way after you have left the room (and get to see which question is posed to you). Start to consider your possible strategies. Maudlin shows by a simple argument that no strategy whatsoever will enable you to reproduce the photons' statistics in the Orsay experiment. The conclusion is that the Bell inequalities do not torpedo just the ambition to build hidden variables models subject to locality; they torpedo locality altogether.

Recently a plethora of new examples, involving more observables or more particles, have been invented, which make this point even more clearly (or at least, in different ways). The simplest of these examples involves again two spin half particles in the special so-called singlet state ρ of the Aspect experiment. We consider the spin observables for the x, y, z directions on each of the two particles; the observables now being normalised to have eigenvalues plus or minus 1. Denote the operators (represented by the tensor products of the Pauli matrices for the one particle with the identity for the other) by X_1, Y_1, Z_1; X_2, Y_2, Z_2. Any observable for particle 1 is compatible with any for particle 2, but for a single particle the operators are pairwise incompatible, in fact they all anticommute: $X_1Y_1 = -Y_1X_1$, etc. We have the identities $X_1Y_1 = iZ_1$, $Y_1Z_1 = iX_1$, $Z_1X_1 = iY_1$ and similarly for particle 2.

Suppose the particles are in the singlet state ρ for which simultaneous measurement of the spins in the same direction of the two particles gives opposite results with probability one; the two possibilities then each having

probability 1/2. Consider the six operators X_1Y_2, Y_1X_2, X_1, Y_2, Y_1, X_2. Suppose a noncontextual hidden variables model existed for measurements of these observables. The measurements can all only be plus or minus one and various functional relations exist within compatible subsets, for instance the two products are products of commuting operators and hence each product together with its two constituents forms a compatible triple. Moreover in the singlet state, measurement of the 'same' spin component of the two particles gives opposite results. We have some more functional relations: $X_1Y_2\,Y_1X_2 = X_1Y_1\,Y_2X_2 = -iZ_1iZ_2 = Z_1Z_2 = iZ_2(-i)Z_1 = X_2Y_2\,Y_1X_1 = X_2Y_1Y_2X_1 = Y_1X_2\,X_1Y_2$. Thus the two products commute with one another and with their own product Z_1Z_2. Suppose measurement of X_1 and Y_2 yields the values x_1 and y_2, so X_1Y_2 yields x_1y_2. Alternatively one could have measured Y_1 and X_2 yielding y_1 and x_2, so that Y_1X_2 yields y_1x_2. Had we measured Y_1 and Y_2 we would have obtained y_1 and y_2 with (with probability one) $y_2 = -y_1$; similarly $x_2 = -x_1$. Therefore $y_1x_2 = x_1y_2$ and their product equals $+1$. However, had we measured X_1Y_2, Y_1X_2, and their product Z_1Z_2 we would have obtained x_1y_2, y_1x_2 and -1: a contradiction.

This example is due to Peres (19??) and it shows that for a particular system in a particular state, a noncontextual hidden variables model is not possible. Mermin (19..) shows how it can be generalised to give a no-go result for noncontextual models for arbitrary state. Though we gave an interpretation in terms of spins of two interacting spin half particles, any quantum system of dimension at least four allows construction of a set of operators having the required properties. Therefore Mermin's extension is an almost general no-go result: it requires $\dim(\mathcal{H}) \geq 4$ rather than the best possible $\dim(\mathcal{H}) \geq 3$.

Here it is: consider the 6×3 array of operators (all having spectrum $\{-1, 1\}$)

$$\begin{array}{ccc} X_1 & Y_2 & X_1Y_2 \\ Y_1 & X_2 & Y_1X_2 \\ X_1 & X_2 & X_1X_2 \\ Y_1 & Y_2 & Y_1Y_2 \\ X_1Y_2 & Y_1X_2 & Z_1Z_2 \\ X_1X_2 & Y_1Y_2 & Z_1Z_2. \end{array} \tag{5.19}$$

We already studied the first five rows. The first two operators in each row commute and their product is the third operator in the row; each row is therefore a commuting triple. The sixth row is also a commuting triple as can be verified by a similar computation as the one we did above; however, now the third operator in that row is *minus* the product of the first two. Note that in the whole table just 9 operators appear, each one twice. The

product of the operators in any row is the square of the last and is the identity operator, with the exception of the last row which yields minus the identity.

Suppose a noncontextual hidden variables model could be constructed for these observables. Then we could assign values plus or minus one to the nine observables in the table such that the product of values in each row is plus one, except the last row, where it is minus one. However this is impossible: the product of all 18 must be plus one since each observable appears twice, while the product of the row products has to be minus one.

The singlet state is a shared eigenstate of the last row of operators. In that state, measurement of each produces the value -1 with certainty. The other rows give three $+1$'s or two -1's and one $+1$, each of the four possibilities with probability $1/4$.

This example cannot be used to provide a no-go result for contextual models subject to locality. The reason for this is as follows. Clearly such a result must assume a state involving interaction of the two particles. Furthermore we may only consider assignments of values to operators, independently of the context, when the corresponding observables can be measured by well separated devices. It makes physical sense to propose that 'observable X_1 gets a certain value, independently of whether X_2 or Y_2 is measured'. However it does not make sense to propose that X_1Y_2 gets a certain value, independently of whether we also measure the pair X_1, Y_2 or the pair Y_1X_2, Z_1Z_2. The simultaneous measurement of the latter triple X_1Y_2, Y_1X_2, Z_1Z_2 requires a highly non-local device to be constructed.

It turns out that only a slightly more complex example does work. Consider now three interacting spin half particles, and consider the following 5×4 array of plus or minus one valued observables. Each row consists of a commuting quadruple; the last operator in each row is the product of the three others except for the final row, where a minus sign is introduced. The product over the whole row yields the identity operator in all rows but the last, where it produces the negative identity. In the whole array ten different observables appear, each one twice:

$$\begin{array}{cccc} X_1 & Y_2 & Y_3 & X_1Y_2Y_3 \\ Y_1 & X_2 & Y_3 & Y_1X_2Y_3 \\ Y_1 & Y_2 & X_3 & Y_1Y_2X_3 \\ X_1 & X_2 & X_3 & X_1X_2X_3 \\ X_1Y_2Y_3 & Y_1X_2Y_3 & Y_1Y_2X_3 & X_1X_2X_3 \end{array} . \qquad (5.20)$$

Suppose a noncontextual hidden variables model was possible. Then we could assign a plus or minus one to each observable in the table, such that the row products are all plus one, except for the last row, which should yield minus

one. This is impossible since the product of all twenty values will always be plus one.

This gives a no-go result for dimension at least 8: not so exciting, but what is nice is that a variant of the example also provides a no-go result for a contextual hidden variables model subject to locality.

For a contextual model subject to locality prepare the three particle system in a shared eigenstate of the bottom row of observables: for instance, simultaneous measurement of them would yield $(+1, +1, +1, -1)$ with certainty. Such a state does exist. In this state, measurement of the first three observables in any one of the first four rows would yield plus or minus ones with product $+1, +1, +1, -1$ according to the row. Now a contextual hidden variables model subject to locality would propose that the value obtained on measuring, for instance, spin in the x-direction on particle 1, does not depend on which spin component is measured on the other two (distant) particles. So the model would lead one to suppose the joint existence of six random variables $X_{1\rho}, Y_{1\rho}, X_{2\rho}, Y_{2\rho}, X_{3\rho}, Y_{3\rho}$ corresponding to the six operators which each appears twice in the 4×3 top left part of the Mermin array. Each represents the result of measuring a certain spin component on a certain particle, independently of which components are measured on the other particles. The joint distribution of these six random variables would be such that the product of triples in each row is, with probability one, $+1, +1, +1, -1$ according to the row. But the product of all the twelve values in this part of the table must be $+1$ since each appears twice: a contradiction.

The no-go results obtained from this table have been described as non-statistical: in the Aspect experiment, it was relative frequencies of various outcomes which, if the sample is large enough, would lead one to statistically reject the hidden variables model. Now we get no-go results by a built-in contradiction in each and every single outcome of the hidden variables model. From a mathematical point of view this is nice; however from an experimental point of view one might well ask what physical experiment could be used to verify the quantum theoretical predictions, and hence empirically reject hidden variables. This question does not seem to have been considered before. Verification of the contextual no-go result would require one to produce interacting triples of particles in this special state, and then check that in the four possible experiments described in the 4×3 top left part of the Mermin array, the products are always $+1, +1, +1, -1$. Empirical verification of the noncontextual result has to be done in a specific state, and then the same experiment as we have just described would suffice.

Chapter 6

Quantum Tomography

6.1 Introduction

Quantum tomography means the reconstruction of the state of an infinite-dimensional quantum system from measurements on that system. The word tomography is used for historical reasons: the first method for this task was based on a mathematical analogy with the problem of reconstructing a two-dimensional image from one-dimensional projections, also known as computer-aided tomography, as it arises in various kinds of medical imaging. We will expand on this connection later. We discuss the simplest possible version of the problem as it comes from quantum optics: a so-called homodyne measurement of a single mode of an electromagnetic field, with perfect detector efficiency. From our perspective one can consider this as the prototype quantum-statistical version of the problem of non-parametric estimation of a probability distribution on the real line, given independent and identically distributed observations from that distribution. In ordinary statistics we realise that different methods are called for when one is interested in different functionals of the distribution, and when one has more or less prior information on the distribution. Thus one might be satisfied with the empirical distribution function as estimator of the distribution function, but would use a kernel estimator or something more sophisticated to estimate the probability density. The appropriate choice of density estimator would depend on prior information about the smoothness of the density; and the optimal rate of convergence will depend on the smoothness too. A similar situation arises in quantum tomography but has so far not been systematically studied.

First of all we introduce some notation. Let $|0\rangle$, $|1\rangle$, $|2\rangle$, ..., denote an orthonormal basis of a separable Hilbert space $\mathcal{H}$. In the context of a homodyne measurement in quantum optics, these represent the state of 0, 1,

2, ... photons in a pulse of light at a particular frequency. We will define various linear operators by specifying their action on the basis elements. These operators are then defined by linearity on the set of finite complex linear combinations of our basis elements, which is a dense subset of $\mathcal{H}$. The question as to what is the maximal domain of definition of each operator is a key issue in the mathematical underpinnings of this subject, but we shall not have to go into details.

Here are the definitions (beware, not all authors use the same ones):

$$\begin{array}{rcll} A^+ |n\rangle &=& \sqrt{n+1}\, |n+1\rangle & \text{Creation} \\ A^- |n\rangle &=& \sqrt{n}\, |n-1\rangle & \text{Annihilation} \\ N |n\rangle &=& n\, |n\rangle & \text{Number} \\ Q &=& (A^- + A^+)/\sqrt{2} & \text{Position} \\ P &=& \frac{1}{i}(A^- - A^+)/\sqrt{2} & \text{Momentum} \\ X_\phi &=& \cos\phi\, Q + \sin\phi\, P & \text{Quadrature at phase } \phi \end{array} \tag{6.1}$$

One should observe that

$$\begin{array}{rcl} N = A^+A^- &=& A^-A^+ - \mathbf{1} = \frac{1}{2}(Q^2 + P^2 - \mathbf{1}) \\ X_0 &=& Q \\ X_{\pi/2} &=& P \\ X_\phi &=& -X_{\phi+\pi}. \end{array} \tag{6.2}$$

Also useful are

$$\begin{array}{rcl} Q + iP &=& \sqrt{2}A^- \\ Q - iP &=& \sqrt{2}A^+ \\ [A^-, A^+] &=& \mathbf{1} \\ [Q, P] &=& i\mathbf{1}. \end{array} \tag{6.3}$$

The names 'position' and 'momentum' come from another application of the same mathematical model—the position and momentum of a particle in a quadratic potential field on the real line, the *quantum simple harmonic oscillator*. For that application, the number operator corresponds to *energy* and the states in the basis are its eigenstates. Our definition of 'quadrature at phase ϕ' differs from the usual definition in quantum optics by a factor of root 2.

It turns out that the operators N, Q, P, and X_ϕ, $\phi \in [0, 2\pi]$ are self-adjoint on their natural (maximal) domains of definition. They therefore each have a spectral decomposition, and in theory each of these operators can be measured on a quantum system in state ρ. The statistical problem which we consider in this chapter is the following: suppose Φ is chosen uniformly at random from $[0, 2\pi]$, and conditional on $\Phi = \phi$, let the random variable X be

the outcome of a measurement of the operator X_ϕ. Given are n independent and identically distributed observations from the joint distribution of (Φ, X). Problem: estimate ρ.

For instance, the probability distribution of the outcome of a measurement of the position operator Q in state ρ is $\mathrm{trace}(\rho\Pi_{[Q\in B]})$ where $\Pi_{[Q\in B]}$ is the projector-valued probability measure associated with Q and $B \subseteq \mathbb{R}$ is a Borel set. For future reference we mention another characterization of this probability distribution. For any bounded Borel function f of Q, the expectation of f of the outcome of measurement of Q is $\mathrm{trace}(\rho f(Q))$. In particular, taking for f the function $\exp(it\cdot)$ we find that the characteristic function of the probability distribution of the outcome of measurement of Q is $\mathrm{trace}(\rho e^{iQt})$. We note that the operator e^{iQt} is defined on all of $\mathcal{H}$ and is unitary. The family of unitary operators e^{iQt} forms a semigroup with generator Q. Commutation relations between the semigroups generated by various of the operators mentioned above, and the trick of characterizing probability distributions by characteristic functions, will provide a means of transfering distributional results from one operator to another.

We will show that the problem is well-posed, in that the state ρ is identified from the joint probability distribution of (Φ, X). In particular, we will show that the matrix elements $\rho_{m,m'} = \langle m' \mid \rho \mid m\rangle$, which determine ρ, can be estimated unbiasedly and with finite variance by sample averages of certain functions of each observation. We will discuss various alternative estimation strategies, and also mention various functionals of the state, and parametrizations (representations) of the state, which could be of interest in practice.

Actually, in the real world the problem will be somewhat harder. A standard homodyne measurement in quantum optics does not produce a measurement of the observable X_ϕ, but a measurement of X_ϕ plus Gaussian noise, with a variance which is related to the efficiency of certain photodetectors. 100% efficiency corresponds to no noise, and this is not experimentally realisable at the moment (and probably never). However this extra complication seems not to essentially change the nature of the statistical problem we are facing, for the kind of detector efficiency met with in practice (say 80% efficient).

We need a number of mathematical tools to solve the problem. On the one hand, we will use an explicit 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 normalized Hermite polynomials times the square root of the error function: the normal density with mean zero and variance half. The observables Q and P become rather easy to describe in this representation. On the other

hand, algebraic results from the theory of representations of groups, provide further relations between the observables X_ϕ, N, Q and P, and also with theory of Fourier transforms.

For the functional analysis, useful references are Biane (1995), Zeidler (1995), Stroock (1999); for the quantum tomography see d'Ariano (1997a,b), Banaszek et al (1999), d'Ariano (1999). Unfortunately definitions of Fourier transform and Hermite polynomial vary tremendously throughout the literature. In particular, as far as the Hermite polynomials are concerned, probabilists base their definition on the standard normal density, while physicists prefer the error function. At first instance the probabilists' choice leads to cleaner formulas, but later on the physicist's choice pays off. On the whole we follow Zeidler (1995), d'Ariano (1997a,b) and the physicists.

6.1.1 Hermite polynomials

The Hermite polynomials $H_n(x)$, $n = 0, 1, 2 \ldots$, are (for us and for the physicists) defined by

$$H_n(x) = e^{x^2}(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}. \tag{6.4}$$

It follows that $H_n(x)$ is an n'th order polynomial with leading term $(2x)^n$. Equvialently, they can be defined from the generating function

$$\exp(2tx - t^2) = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x). \tag{6.5}$$

It turns out that they can also be obtained from the simple polynomials $(2x)^2$, $n = 0, 1, 2, \ldots$ by Gramm-Schmidt orthogonalisation with respect to the normal density with mean zero and variance half, also known as the error function $\mathrm{erf}(x) = (1/\sqrt{\pi})\exp(-x^2)$. The Hermite polynomials are orthogonal with respect to the error function, but not normalized. It turns out that if X is normal with mean zero and variance half, then $\mathrm{E}(H_n(X)^2) = 2^n n!$. By dividing by the square root of $2^n n!$ and multiplying by the square root of the error function we therefore produce an orthonormal sequence u_n in the space of square integrable functions with respect to Lebesgue measure on the line:

$$u_n(x) = \sqrt{\frac{\mathrm{erf}(x)}{2^n n!}} H_n(x). \tag{6.6}$$

This sequence is not only orthonormal but complete—it forms a basis of $L^2_{\mathbb{C}}(\mathbb{R})$. This means that one can set up a Hilbert space isomorphism, or

if you prefer a unitary equivalence, between our abstract space $\mathcal{H}$ and the concrete space $L^2_{\mathbb{C}}(\mathbb{R})$. Now from the generating function one can derive two key recurrence formulas for the the functions u_n

$$\sqrt{2}xu_n(x) = \sqrt{n+1}u_{n+1}(x) + \sqrt{n}u_{n-1}x, \tag{6.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}u_n(x) = \sqrt{2}\sqrt{n}u_{n-1}(x) - xu_n(x). \tag{6.8}$$

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

$$\begin{array}{rcl} Q = \sqrt{2}(A^- + A^+)/\sqrt{2} & \longleftrightarrow & x \times \cdot \\ P = \frac{1}{i}(A^- - A^+)/\sqrt{2} & \longleftrightarrow & \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x} \\ 2N + \mathbf{1} = Q^2 + P^2 & \longleftrightarrow & (x^2 - \frac{\mathrm{d}^2}{\mathrm{d}x^2}) \end{array} \tag{6.9}$$

In Chapter 3 (infinite dimensional theory) we showed that the operator Q 'multiplication by X' on the space $L^2_{\mathbb{C}}(\mathbb{R})$ is associated with the projector-valued measure $\Pi_{[Q\in B]}$ equal to 'multiplication by $1_B(x)$', the indicator function for Borel set $B \subseteq \mathcal{R}$. In other words, the subspace $[Q \in B]$ is the subspace of L^2 functions with support in B. Thus for a pure state $|\psi\rangle \in \mathcal{H}$ represented by the wave function $\psi(x) \in L^2_{\mathbb{C}}(\mathbb{R})$, the probability that a measurement of Q takes a value in B is equal to $\|1_B\psi\|^2 = \int_B |\psi(x)|^2\mathrm{d}x$, thus the outcome of the measurement has probability density $|\psi(x)|^2$. We can write $|\psi\rangle = \sum_m c_m |m\rangle$ where $c_m = \langle m \mid \psi\rangle = \int u_m(x)\psi(x)\mathrm{d}x$, and the infinite sum converges in L^2 sense. If the sum actually converges pointwise, the probability density can also be written as

$$\sum_m |c_m u_m(x)|^2 = \sum_m \sum_{m'} \overline{c}_m c_{m'} u_m(x) u_{m'}(x). \tag{6.10}$$

In particular, if the state $|\psi\rangle = |0\rangle$, then the distribution of the outcome of a measurement of Q is the normal density with mean zero and variance half.

Suppose now the state is mixed, and obtained from pure states $|\psi_k\rangle = \sum_m c_{k,m}|m\rangle$ by mixing with respect to the probabilities p_k. If we denote the mixed state by ρ it follows that $\rho_{m,m'} = \sum_k p_k \overline{c_{k,m}} c_{k,m'}$. Therefore, by mixing our previous expression over the probabilities p_k, we might hope that the probability density of a measurement of Q on the state ρ is

$$\sum_m \sum_{m'} \rho_{m,m'} u_m(x) u_{m'}(x). \tag{6.11}$$

As it stands this expression is not necessarily meaningful, since our derivation assumes pointwise convergence when we only are guaranteed L^2 limits. However, if one truncates the double infinite sum at a finite level M, one obtains a sequence of functions with total mass $\sum_{m=1}^{M} \rho_{m,m}$, yielding a sequence of sub-probability distributions which converge as $M \to \infty$ in the sense of convergence in distribution to the distribution of the outcome of the measurement.

Shortly we will be able to generate from this result, the probability distribution of a measurement of any of the operators X_ϕ in an arbitrary mixed state ρ.

The observable N has eigenvalues n and eigenfunctions $|n\rangle$. Hence measurement of N on the pure state ψ results in the outcome m with probability $|c_m|^2$ where as before $c_m = \langle m \mid \psi \rangle = \int u_m(x)\psi(x)\mathrm{d}x$. On a mixed state ρ the outcome is m with probability $\langle m \mid \rho \mid m \rangle = \rho_{m,m}$.

6.1.2 Generalized Fourier transforms

The next item of functional analysis we need is the beautiful relation between Fourier transforms and Hermite polynomials. Again there are many different definitions in the literature. Recall that $u_n(x)$ denotes the n'th normalized Hermite polynomial, multiplied by the square root of the error function. Then it turns out that the Fourier transform of u_n is equal to $(-i)^n$ times itself, and the inverse Fourier transform is i^n times itself:

$$\frac{1}{\sqrt{2\pi}} \int_{x=-\infty}^{\infty} e^{-itx} u_n(x)\mathrm{d}x = (-i)^n u_n(t), \tag{6.12}$$

$$\frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{\infty} e^{itx} u_n(t)\mathrm{d}x = i^n u_n(x). \tag{6.13}$$

Define the operator F by $F|n\rangle = (-i)^n |n\rangle$, and $F^* |n\rangle = i^n |n\rangle$. Thus $F = \exp(-i(\pi/2)N)$ and $F^* = \exp(i(\pi/2)N)$ are unitary operators, one another's adjoints, extending the Fourier transform and its inverse in a continuous way, in the L^2 sense, from the basis elements $u_n(x)$ of $L^2_{\mathbb{C}}(\mathbb{R})$ to all of that space. Now the Fourier transform of $1/i$ times the derivative of a smooth function is t times the transform of the function itself. Thus in terms of the operators Q (position) and P (momentum), we have $FQ = PF$ or $P = F^*QF$. The same relation holds for functions of Q and P, in particular $1_B(P) = \Pi_{[P \in B]} = F^* 1_B(Q) F$ for any Borel set B. Now the probability that a measurement of P on a pure state $|\psi\rangle$ takes a value in the Borel set B equals $\| 1_B(P) |\psi\rangle \|^2 = \| F^* 1_B(Q) F |\psi\rangle \|^2 = \| 1_B(Q) F |\psi\rangle \|^2 = \int_B |F|\psi\rangle(x)|^2 \mathrm{d}x$.

Thus the probability density of the outcome of a measurement of the momentum operator P on a pure state ψ is equal to $|(F\psi)(x)|^2$, with F the Fourier transform (extended to arbitrary L^2 functions, and defined modulo L^2 equivalence). In particular, in the state $|0\rangle$ the density is normal, mean zero, variance half. For mixed states we can make similar remarks.

These results can be generalized even further. Let $z = re^{i\phi}$ be a complex number and define the unitary operator, called a Weyl operator (again, other definitions are possible), $W_z = \exp(irX_\phi)$ where $X_\phi = \cos\phi Q + \sin\phi P$ is the quadrature at phase ϕ defined at the outset. Now it turns out that $e^{i\theta N} W_z e^{-i\theta N} = W_{e^{i\theta}z}$, or in terms of X_ϕ, $e^{i\theta N} e^{itX_\phi} e^{-i\theta N} = e^{itX_{\phi+\theta}}$.

First we show how this relation allows us to rederive the relations between the distributions of measurements of Q and of P we found before. With $\theta = \pi/2$ and $\phi = 0$ we recover the relation between position Q, and momentum P and Fourier transform F, $F^* e^{itQ} F = e^{i(\pi/2)N} e^{itQ} e^{-i(\pi/2)N} = e^{itP}$. In the mixed state ρ the characteristic function of a measurement of the position operator Q is $\mathrm{trace}(\rho e^{itQ})$. Similarly, that of a measurement of the momentum operator is

$$\begin{aligned} \mathrm{trace}(\rho e^{itP}) &= \mathrm{trace}(\rho e^{i(\pi/2)N} e^{itQ} e^{i(-\pi/2)N}) \\ &= \mathrm{trace}((e^{-i(\pi/2)N} \rho e^{i(\pi/2)N}) e^{itQ}) \\ &= \mathrm{trace}((F\rho F^*) e^{itQ}). \end{aligned} \tag{6.14}$$

If ρ is a pure state $|\psi\rangle\langle\psi|$ then the final result is equal to

$$\mathrm{trace}(|F\psi\rangle\langle F\psi| e^{itQ}) = \langle F\psi \mid e^{itQ} \mid F\psi \rangle . \tag{6.15}$$

In other words, the distribution of a measurement of P on state $|\psi\rangle$ is the same as that of a measurement of Q on the Fourier transform of ψ.

For the observable X_ϕ, since $e^{i\theta N} e^{itX_\phi} e^{-i\theta N} = e^{itX_{\phi+\theta}}$ and $X_0 = Q$, we have $e^{i\phi N} e^{itQ} e^{-i\phi N} = e^{itX_\phi}$. For the pure state case, if $\psi = \sum c_m |m\rangle$ then the distribution of a measurement of X_ϕ in state $|\psi\rangle$ is the same as that of a measurement of Q in the state $e^{-i\phi N}|\psi\rangle = \sum e^{-mi\phi} c_m |m\rangle$. Thus from (6.10), the density of this distribution can be formally written as

$$\sum_m \sum_{m'} \overline{c}_m c_{m'} e^{i(m-m')\phi} u_m(x) u_{m'}(x). \tag{6.16}$$

The word formally means that the double summation does not necessarily converge pointwise. To be precise, the density is obtained via a limiting operation in the L^2 sense. In the mixed case, generalizing (6.11), we have formally the probability density

$$\sum_m \sum_{m'} \rho_{m,m'} e^{i(m-m')\phi} u_m(x) u_{m'}(x). \tag{6.17}$$

These expressions are both literally correct when only a finite number of terms are not zero. In general, the truncated sums provide approximations to the actually density in the sense of convergence in distribution.

6.2 Tomography

Can one recover the state ρ from the probability distribution of measurements of all X_ϕ? The answer to this question is yes. A most elegant way to see this follows from the following identity of d'Ariano,

References

A. Aspect, 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.

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

O.E. Barndorff-Nielsen and R.D. Gill (1998). *An example of nonattainability of expected quantum information.* Preprint `quant-ph/9802063, http://xxx.lanl.gov`

O.E. Barndorff-Nielsen and R.D. Gill (2000). *Fisher information in quantum statistics.* Preprint. `http://www.math.uu.nl/people/gill/Preprints/unattain.ps.gz`

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

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

J.S. Bell (1966). On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447–452.

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

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

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

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

A. Cabello, J.M. Estebaranz and G. García-Alcaine (1996). Bell-Kochen-Specker theorem: a proof with 18 vectors. *J. Phys. Lett. A* **212**, 183–187.

R. Cooke, M. Keane and W. Moran (1985). An elementary proof of Gleason's theorem. *Math. Proc. Camb. Phil. Soc.* **98**, 117–128.

J. Dorling J. (1992). *A simple logician's guide to the quantum puzzle and to quantum logic's putative solution*, preprint, Univ. Amsterdam.

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

A. Fine (1982). Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **48**, 291–295.

B.C. van Fraassen (1991). *Quantum Mechanics: An Empiricist View*, Oxford University Press.

A. Galindo (1976). Another proof of the Kochen-Specker theorem. *Algunas cuestiones de fisica teorica 1.975* (In memoriam Angel Esteve), 3–9, GIFT, Zaragoza.

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

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

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

R.D. Gill (2000) Asymptotics in quantum statistics. To appear in *State of the Art in Mathematical Statistics and Probability*, ed. A.W. van der Vaart, M. van de Gunst, C.A.J. Klaassen, IMS monographs. Preprint `http://www.math.uu.nl/people/gill/Preprints/paper.ps.gz`

S. Goldstein (1998). Quantum mechanics without observers. *Physics Today*, March, April 1998; letters to the editor, it Physics Today, February 1999.

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

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

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

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

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

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

A.N. Kolmogorov (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin.

S. Kochen and E.P. Specker (1967). The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **17** 59–87.

K. Kraus (1983). *States, Effects, and Operations.* Springer Lecture Notes in Physics **190**.

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

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

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

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

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

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

J. von Neumann (1932). *Mathematical Foundations of Quantum Mechanics.* English Translation, 1955, Princeton University Press.

M. Ozawa (1984). Quantum measuring processes of continuous observables. *I. Math. Phys.* **25**, 79–87.

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

M. Ozawa (1988). Measuring processes and repeatability hypothesis. pp. 412–421 in: *Probability Theory and Mathematical Statistics: Proc. Vth Japan-USSR Symp., Kyoto, 1986*, ed. S. Watanabe and Yu.V. Prokhorov, Springer Lecture Notes in Mathematics **1299**.

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

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

A. Peres (1991). Two simple proofs of the Kochen-Specker theorem. *J. Phys. A: Math. Gen.* **24** (Letter to the editor), L175–L178.

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

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

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

M. Reed and B. Simon (1979). *Methods of Modern Mathematical Physics (Vol I).* Academic Press, San Diego, Ca.

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

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

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

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

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

H.P. Yuen and M. Lax (1973). Multiple-parameter quantum estimation and measurement of nonselfadjoint observables. *Trans. IEEE* **IT-19**, 740–750.