

Conceptual variables, quantum theory, and statistical inference theory

Inge S. Helland, Department of Mathematics, University of Oslo,
P.O.Box 1053 Blindern, N-0316 Oslo, Norway
email: ingeh@math.uio.no

December 9, 2019

Abstract

A new approach towards quantum theory is proposed in this paper. The basis is taken to be conceptual variables, variables that may be accessible or inaccessible, i.e., it may be possible or impossible to assign numerical values to them. Group actions are defined on these variables. It is shown that the whole theory may be realized in a finite-dimensional complex Hilbert space when the variables are discrete. The Born formula is derived under weak assumptions. Connections to ordinary statistical inference are discussed, also through an example and through a brief discussion of quantum measurement theory. The interpretation of quantum states (or eigenvector spaces) implied by this approach is as focused questions to nature together with sharp answers to those questions. Resolutions of the identity are connected to the questions themselves, which may be complementary in the sense defined by Bohr.

1 Introduction

One purpose of this article is to discuss connections between some of the latest developments in quantum foundation and aspects of the latest developments in the foundation of statistical inference theory. The goal is to seek links between these two scientific areas, but also to point at differences. For this we need a common language, here expressed by conceptual variables. Another major purpose of the article is to give an introduction to the approach towards formal quantum theory through such variables.

In this article, as in [1], [9] and [11], a conceptual variable is defined as any variable defined by a person or by a group of communicating persons. A conceptual variable is called accessible if it, through an experiment or in other ways, is possible to assign numerical values to it; otherwise it is called inaccessible. For example, by Heisenberg's inequality the vector (position, momentum) for a particle is inaccessible. These notions are particularly interesting in the case of an epistemic process, a process to achieve knowledge. Epistemic processes are discussed in detail in [1]. Two basic such processes are statistical experiments and quantum measurements, but there are also other types of epistemic processes.

As will be indicated below, this point of departure together with symmetry assumptions has implications both for the foundation for and the interpretation of quantum theory. First, it implies under certain conditions the Hilbert space formulation, including introductions of operators for accessible variables and state vectors. Secondly it makes the following interpretation natural: State vectors, or more generally eigenvector spaces, are in correspondence with focused questions to nature together with sharp answers to these questions.

Taking this loosely defined notion of conceptual variables as a basis, and then postponing the problem of making this notion concrete, allows us largely to stay agnostic concerning some of the debates on the interpretation of quantum theory and the debates between the different schools in statistical inference theory. However, the epistemic view in quantum theory turns out to be useful, and an ideal observer may act as some kind of Bayesian in this theory.

It is crucial in our context here that we assume that there are defined transformation groups on the spaces of conceptual variables that are discussed. This may be the group of all automorphisms on the space in which the conceptual variable varies, or it may be a concrete subgroup of this group.

In this paper I discuss such situations more closely with a focus on the more mathematical aspects of the notions and situations described above. I assume the existence of a concrete (physical) situation, and that there is a space Ω_ϕ of the inaccessible conceptual variable ϕ with a group K acting on this space. There is an e-variable, an accessible conceptual variable, θ defined, a function on Ω_ϕ . This θ varies on a space Ω_θ , and the group K may induce a transformation group G on Ω_θ . A very simple situation is when ϕ is a spin vector, and θ is a spin component in a given direction, but this paper focuses on more general situations. Also, in the simple spin situation the natural group K for the spin vector does not induce directly groups on the component, see a discussion in Section 6 below.

When there are more potential e-variables, I will denote this by a superscript a : θ^a and $G^a = \{g^a\}$. Note that in Chapter 4 of [1] I used different notations for the groups: G and \tilde{G}^a for what is here called K and G^a . The group G^a which was referred to there, is here the subgroup of K corresponding to G^a , denoted by H^a below. Both here and in [1] I will use the word ‘group’ as synonymous to ‘group action’ or transformation group on some set, not as an abstract group.

The article has some overlap with the paper [9], but the perspective here is wider.

In Section 2 I introduce some basic group theory that is needed in the paper. Then in Sections 3-9 I formulate a completely new approach to the foundation of quantum theory. The basis is conceptual variables with group actions defined on them. From this, the ordinary Hilbert space formalism is derived, and it is indicated how operators corresponding to accessible physical variables may be derived. In the case of discrete variables, taking d different values, it is shown that under weak conditions all operators may be constructed relative to the Hilbert space \mathbb{C}^d , where \mathbb{C} is the complex field. The Born formula is derived by using two assumptions: 1) A focused likelihood principle, derived from the ordinary likelihood principle of statistics; 2) An assumption of a perfectly rational ideal observer. In Section 10 various approaches to ordinary statistical inference are discussed, and in Section 11 an example is given which lies in the border area between quantum theory and Bayesian statistical inference. Section

12 discusses briefly quantum measurement theory, while a final discussion is given in Sections 13-14.

2 Some basic group theory

Let ϕ be an inaccessible conceptual variable varying in a space Ω_Φ . It is a basic philosophy of the present paper that I always regard groups as group actions or transformations, acting on some space.

Starting with Ω_Φ and a group K acting on Ω_Φ , let $\theta(\cdot)$ be an accessible function on Ω_Φ , and let Ω_Θ be the range of this function.

As mentioned in the Introduction, I regard ‘accessible’ and ‘inaccessible’ as primitive notions. Ω_Θ and Ω_Φ are equipped with topologies, and all functions are assumed to be Borel-measurable.

Definition 1. *The e-variable θ is maximally accessible if the following holds: If θ can be written as $\theta = f(\psi)$ for a function f that is not one-to-one, the conceptual variable ψ is not accessible. In other words: θ is maximal under the partial ordering defined by $\alpha \leq \beta$ iff $\alpha = f(\beta)$ for some function f .*

Note that this partial ordering is consistent with accessibility: If β is accessible and $\alpha = f(\beta)$, then α is accessible. Also, ϕ is an upper bound under this partial ordering. The existence of maximally accessible conceptual variables follows then from Zorn’s lemma.

Definition 2. *The accessible variable θ is called permissible if the following holds: $\theta(\phi_1) = \theta(\phi_2)$ implies $\theta(k\phi_1) = \theta(k\phi_2)$ for all $k \in K$.*

With respect to parameters and subparameters along with their estimation, the concept of permissibility is discussed in some details in Chapter 3 in [2]. The main conclusion, which also can be generalized to this setting, is that under the assumption of permissibility one can define a group G of actions on Ω_Θ such that

$$(g\theta)(\phi) := \theta(k\phi); k \in K. \quad (1)$$

Herein I use different notations for the group actions g on Ω_Θ and the group actions k on Ω_Φ ; in contrast, the same symbol g was used in [2]. The background for that is

Lemma 1. *Assume that θ is a permissible variable. The function from K to G defined by (1) is then a group homomorphism.*

Proof. Let k_i be mapped upon g_i by (1) for $i = 1, 2$. Then, for all $\phi \in \Omega_\Phi$ we have $(g_i\theta)(\phi) = \theta(k_i\phi)$. Assume that $k_2\phi$ is mapped to $\theta' = \theta(k_2\phi) = (g_2\theta)(\phi)$. Then also $\theta(k_1k_2\phi) = (g_1\theta')(k_2\phi) = (g\theta)(\phi)$ for some g . Thus $(g_1(g_2\theta))(\phi) = (g\theta)(\phi)$ for all ϕ , and since the mapping is permissible, we must have $g = g_1g_2$.

It is important to define left and right invariant measures, both on the groups and on the spaces of conceptual variables. In the mathematical literature, see for instance

[3,4], Haar measures on the groups are defined (assuming locally compact groups). Right (μ_G) and left (ν_G) Haar measures on the group G satisfy

$$\mu_G(Dg) = \mu_G(D), \text{ and } \nu_G(gD) = \nu_G(D)$$

for $g \in G$ and $D \subset G$, respectively.

Next define the corresponding measures on Ω_θ . As is commonly done, I assume that the group operations $(g_1, g_2) \mapsto g_1 g_2$, $(g_1, g_2) \mapsto g_2 g_1$ and $g \mapsto g^{-1}$ are continuous. Furthermore, I will assume that the action $(g, \theta) \mapsto g\theta$ is continuous.

As discussed in Wijsman [5], an additional condition is that every inverse image of compact sets under the function $(g, \theta) \mapsto (g\theta, \theta)$ should be compact. A continuous action by a group G on a space Ω_θ satisfying this condition is called *proper*. This technical condition turns out to have useful properties and is assumed throughout this paper. When the group action is proper, the orbits of the group can be proved to be closed sets relative to the topology of Ω_θ .

The connection between ν_G defined on G and the corresponding left invariant measure ν defined on Ω_θ is relatively simple: If for some fixed value θ_0 of the conceptual variable the function β on G is defined by $\beta : g \mapsto g\theta_0$, then $\nu(E) = \nu_G(\beta^{-1}(E))$. This connection between ν_G and ν can also be written $\nu_G(dg) = d\nu(g\theta_0)$, so that $d\nu(hg\phi_0) = d\nu(g\phi_0)$ for all $h, g \in G$.

The following result, originally due to Weil is proved in [5]; for more details on the right-invariant case, see also [2].

Theorem 1. *The left-invariant measure ν on Ω_θ exists if the action of G on Ω_θ is proper and the group is locally compact.*

Note that ν can be seen as an induced measure on each orbit of G on Ω_θ , and it can be arbitrarily normalized on each orbit. ν is finite on a given orbit if and only if the orbit is compact. In particular, ν can be defined as a probability measure on Ω_θ if and only if all orbits of Ω_θ are compact. Furthermore, ν is unique only if the group action is transitive.

In a corresponding fashion, a right invariant measure can be defined on Ω_θ . This measure satisfies $d\mu(gh\phi_0) = d\mu(g\phi_0)$ for all $g, h \in G$. In many cases the left invariant measure and the right invariant measure are equal.

3 Some quantum theory and a discussion on interpretation

It is assumed that the reader is familiar with the ordinary quantum theory for discrete observables: For each physical system there is a Hilbert space \mathcal{H} , and each observable θ on the system is associated with a self-adjoint operator A . The possible values of θ are the eigenvalues of A . The states of the system are the unit vectors $|u\rangle$ of \mathcal{H} , and when $|u\rangle$ is an eigenvector of A corresponding to an eigenvalue u_k , then $\theta = u_k$ with certainty. If necessary, the minimum amount of theory is for instance given in [29].

The purpose of the papers [9,11] has been to rederive this formal theory from assumptions about conceptual variables. On the following sections the discussion of [9] is included for completeness. Results from the more recent [11] will also be needed.

As is well known, the probabilities of quantum theory are calculated by the Born rule. One version of this rule is: Assume that the physical system has been prepared in the state $|u\rangle$, corresponding to $\theta^a = u_k$ for some observable θ^a , and let the aim be to measure another observable θ^b . Then

$$P(\theta^b = v_j) = \langle u | \Pi_j | u \rangle, \quad (2)$$

where Π_j is the projector upon the eigenspace corresponding to the eigenvalue $\theta^b = v_j$. The Born rule is derived from a reasonable set of assumptions in [1]; see a discussion below.

The reader is also probably aware of the fact that there exist several interpretations of quantum theory, and that the discussions between the supporters of the different interpretations is still going on. During the recent years there has been held a long range of international conferences on the foundation of quantum mechanics. A great number of interpretations have been proposed; some of them look very peculiar to the laymen. For instance, the many worlds interpretation assumes that there exist millions or billions of parallel worlds, and that a new world appears every time when one performs a measurement.

On two of these conferences recently there was taken an opinion poll among the participants [30,31]. It turned out to be an astonishing disagreement on many fundamental and fairly simple questions. One of these questions was: Is the quantum mechanics a description of the objective world, or is it only a description of how we obtain knowledge about reality? The first of these descriptions is called ontological, the second epistemic. Up to now most physicists have supported the ontological or realistic interpretation of quantum mechanics, but versions of the epistemic interpretation have received a fresh impetus during the recent years.

I look upon my book 'Epistemic Processes' [1] as a contribution to this debate. An epistemic process can denote any process to achieve knowledge. It can be a statistical investigation or a quantum mechanical measurement, but it can also be a simpler process. The book starts with an informal interpretation of quantum states, which in the traditional theory has a very abstract definition. In my opinion, a quantum state always has a connection to a focused question and a sharp answer to this question.

A related interpretation is QBism, or quantum Bayesianism, see Fuchs [27] and von Baeyer [28]. The predictions of quantum mechanics involve probabilities, and a QBist interpret these as purely subjective probabilities, attached to a concrete observer. Many elements in QBism represent something completely new in relation to classical physical theory, in relation to many people's conception about science in general and also to earlier interpretations of quantum mechanics. The essential thing is that the observer plays a role that can not be eliminated. The single person's comprehension of reality can differ from person to person, at least at a given point of time, and this is in principle all that can be said.

Such an understanding can in my opinion be made valid for very many aspects of reality. We humans can have a tendency to experience reality differently. Partly, this

can be explained by the fact that we give different meaning to the concepts we use. Or we can have different contexts for our choices. An important aspect is that we focus differently.

This statement, that the observer (or a group of communicating observers) in a way can be said to determine aspects of reality, I have tried to give a precise background for in the newer articles [9,11], a background which also is reproduced here.

This background, as I see it, can be expressed through what I call conceptual variables, variables defined by a person or by a group of communicating persons. Such variables can be accessible, that is, it is possible to find values for them, for instance by doing experiments. But some variables are so extensive that it is impossible to find values for them. In a physical situation this can be a vector (position, velocity) for a particle. According to Heisenberg's uncertainty relation it is impossible at the same time to determine position and velocity. Such variables are called inaccessible. In a given situation we can have a fundamental inaccessible variable, and we have the choice between focusing on different accessible (mathematical) functions of this variable. In the physical situation that I sketched, the observer can focus on position or velocity.

By using group theory and group representation theory, I aim at studying such a situation mathematically, and it seems to appear that an essential part of the quantum formulation can be derived under very weak conditions. My opinion is that what we can develop from such results can be crucial for our views on the world around us. Empirically, the quantum formalism has turned out to give a very extensive description of our world as we know it, in physical situations in microcosmos an all-embracing description.

In decision situations and in cognitive modeling it has also been fruitful to look at a quantum description, see [37,38]. In a decision situation the decision variable may be so extensive that it is impossible for the person in question to make a decision; this variable is then called inaccessible. The person can then focus on a simpler, accessible, decision variable, in such a way that it is possible to make a partial decision.

These investigations support the hypothesis that the quantum states in physics are in correspondence with focused questions to nature together with sharp answers to these questions. Sometimes it can be necessary to have an epistemic way to relate to the world; it can be useful to seek knowledge. We can get knowledge on certain issues by focusing on certain questions to the world around us, and our knowledge depends on the answers we obtain to these questions. And this is all we can achieve.

Following such an opinion, it can be argued that for certain phenomena there exists no other reality than this (subjective) attached to each single person. This statement must be made precise to be understood in the correct way. First, it is connected to an ideal observer. Secondly, groups of observers that communicate, can go in and act as one observer when a concrete measurement is focused on. When all potential observers agree on the answer to a measurement, this measurement must represent an objective property of reality. Thus the objective world exists; it is the state attached to certain aspects of the world that in certain cases must be connected to an observer (or to several communicating observers).

Nevertheless, these are aspects of physics – and science – which can be surprising for many people, but in my opinion such viewpoints may be necessary, not only in

physics, but also in many other areas of life.

In fact, this can in my view be made valid for very many aspects of reality. We humans have a tendency to experience reality differently. This can be due to the fact that we give different meanings to the concepts we use. Or we have different contexts for our valuations. An important aspect is that we focus differently.

Here is one remark concerning QBism, which can be said to represent a variant of this view: Subjective Bayes-probabilities have also been in fashion among groups of statisticians. In my opinion it can be very fruitful to look for analogies between statistical inference theory and quantum mechanics, but then one must look more broadly upon statistics and statistical inference theory, not only focus on subjective Bayesianism. This is only one of several philosophies that can form a basis for statistics as a science. Studying connections between these philosophies, is an active research area today; some results are given in Section 10 below.

4 Operators and quantization

In the quantum-mechanical context defined in [1], θ is an accessible variable, and one should be able to introduce an operator associated with θ . The following discussion which is partly inspired by [12], considers an irreducible unitary representation of G on a complex Hilbert space \mathcal{H} .

4.1 A brief discussion of group representation theory

A group representation of G is a continuous homomorphism from G to the group of invertible linear operators V on some vector space \mathcal{H} :

$$V(g_1 g_2) = V(g_1) V(g_2). \quad (3)$$

It is also required that $V(e) = I$, where I is the identity, and e is the unit element of G . This assures that the inverse exists: $V(g)^{-1} = V(g^{-1})$. The representation is unitary if the operators are unitary ($V(g)^\dagger V(g) = I$). If the vector space is finite-dimensional, we have a representation $D(V)$ on the square, invertible matrices. For any representation V and any fixed invertible operator U on the vector space, we can define a new equivalent representation as $W(g) = UV(g)U^{-1}$. One can prove that two equivalent unitary representations are unitarily equivalent; thus U can be chosen as a unitary operator.

A subspace \mathcal{H}_1 of \mathcal{H} is called invariant with respect to the representation V if $u \in \mathcal{H}_1$ implies $V(g)u \in \mathcal{H}_1$ for all $g \in G$. The null-space $\{0\}$ and the whole space \mathcal{H} are trivially invariant; other invariant subspaces are called proper. A group representation V of a group G in \mathcal{H} is called irreducible if it has no proper invariant subspace. A representation is said to be fully reducible if it can be expressed as a direct sum of irreducible subrepresentations. A finite-dimensional unitary representation of any group is fully reducible. In terms of a matrix representation, this means that we can always find a $W(g) = UV(g)U^{-1}$ such that $D(W)$ is of minimal block diagonal form. Each one of these blocks represents an irreducible representation, and they are all one-dimensional if and only if G is Abelian. The blocks may be seen as operators on subspaces of

the original vector space, i.e., the irreducible subspaces. The blocks are important in studying the structure of the group.

A useful result is Schur's Lemma:

Let V_1 and V_2 be two irreducible representations of a group G ; V_1 on the space \mathcal{H}_1 and V_2 on the space \mathcal{H}_2 . Suppose that there exists a linear map T from \mathcal{H}_1 to \mathcal{H}_2 such that

$$V_2(g)T(v) = T(V_1(g)v) \quad (4)$$

for all $g \in G$ and $v \in \mathcal{H}_1$.

Then either T is zero or it is a linear isomorphism. Furthermore, if $\mathcal{H}_1 = \mathcal{H}_2$, then $T = \lambda I$ for some complex number λ .

Let ν be the left-invariant measure of the space Ω_Θ induced by the group G , and consider the Hilbert space $\mathcal{H} = L^2(\Omega_\Theta, \nu)$. Then the left-regular representation of G on \mathcal{H} is defined by $U^L(g)f(\phi) = f(g^{-1}\phi)$. This representation always exists, and it can be shown to be unitary, see [7].

If V is an arbitrary representation of a compact group G in \mathcal{H} , then there exists in \mathcal{H} a new scalar product defining a norm equivalent to the initial one, relative to which V is a unitary representation of G .

For references to some of the vast literature on group representation theory, see Appendix A.2.4 in [2].

4.2 A resolution of the identity

In the following I assume that the group G has representations that give square-integrable coherent state systems (see page 43 of [12]). For instance this is the case for all representations of compact semisimple groups, representations of discrete series for real semisimple groups, and some representations of solvable Lie groups.

Let G be an arbitrary such group, and let $V(g)$ be one of its unitary irreducible representations acting on the Hilbert space \mathcal{H} . Assume that G is acting transitively on a space Ω_Θ , and fix $\theta_0 \in \Omega_\Theta$. Then every $\theta \in \Omega_\Theta$ can be written as $\theta = g\theta_0$ for some $g \in G$. I also assume that the isotropy group of G is trivial. Then this establishes a one-to-one correspondence between G and Ω_Θ . Note the similarity with the assumptions made in certain approaches to statistical inference, see Subsection 10.2 below.

Also, fix a vector $|\theta_0\rangle \in \mathcal{H}$, and define the coherent states $|\theta\rangle = |\theta(g)\rangle = V(g)|\theta_0\rangle$. With ν being the left invariant measure on Ω_Θ , introduce the operator

$$T = \int |\theta(g)\rangle \langle \theta(g)| d\nu(g\theta_0). \quad (5)$$

Note that the measure here is over Ω_Θ , but the elements are parametrized by G . T is assumed to be a finite operator.

Lemma 2. *T commutes with every $V(h); h \in G$.*

Proof. $V(h)T =$

$$\begin{aligned} \int V(h)|\theta(g)\rangle\langle\theta(g)|d\nu(g\theta_0) &= \int |\theta(hg)\rangle\langle\theta(g)|d\nu(g\theta_0) \\ &= \int |\theta(r)\rangle\langle\theta(h^{-1}r)|d\nu(h^{-1}r\theta_0). \end{aligned}$$

Since $|\theta(h^{-1}r)\rangle = V(h^{-1}r)|\theta_0\rangle = V(h^{-1})V(r)|\theta_0\rangle = V(h)^\dagger|\theta(r)\rangle$, we have $\langle\theta(h^{-1}r)| = \langle\theta(r)|V(h)$, and since the measure ν is left-invariant, it follows that $V(h)T = TV(h)$.

From the above and Schur's Lemma it follows that $T = \lambda I$ for some λ . Since T by construction only can have positive eigenvalues, we must have $\lambda > 0$. Defining the measure $d\mu(\theta) = \lambda^{-1}d\nu(\theta)$ we therefore have the important resolution of the identity

$$\int |\theta\rangle\langle\theta|d\mu(\theta) = I. \quad (6)$$

For a more elaborate similar construction taking into account the so-called isotropy subgroup, see Chapter 2 of [12]. In [9] a corresponding resolution of the identity is derived for states defined through representations of the group K acting on Ω_Φ .

4.3 Quantum operators

Let Ω_Φ be the space of inaccessible conceptual variables, and let $\theta(\phi)$ be an accessible e-variable that is assumed to be permissible with respect to the group K . Let G be the resulting transformation on Ω_Θ , and use the result of the previous subsection.

In general, an operator corresponding to θ may be defined by

$$A = A^\theta = \int \theta|\theta\rangle\langle\theta|d\mu(\theta). \quad (7)$$

A is defined on a domain $D(A)$ of vectors $|v\rangle \in \mathcal{H}$ where the integral defining $\langle v|A|v\rangle$ converges.

This mapping from an e-variable θ to an operator A has the following properties:

- (i) If $\theta = 1$, then $A = I$.
- (ii) If θ is real-valued, then A is symmetric (for a definition of this concept for operators and its relationship to self-adjointness, see [14].)
- (iii) The change of basis through a unitary transformation is straightforward.

For further important properties, we need some more theory. First we consider the situation where we do not longer assume that the mapping from ϕ to $\theta(\phi)$ is permissible with respect to K , but just start with any group G acting on θ .

Theorem 3. *Let H be the subgroup of K consisting of any transformation h such that $\theta(h\phi) = g\theta(\phi)$ for some $g \in G$. Then H is the maximal group under which the e-variable θ is permissible.*

Proof. Let $\theta(\phi_1) = \theta(\phi_2)$ for all $\theta \in \Theta$. Then for $h \in H$ we have $\theta(h\phi_1) = g\theta(\phi_1) = g\theta(\phi_2) = \theta(h\phi_2)$, thus θ is permissible under the group H . For a larger group, this argument does not hold. \square

Next look at the mapping from θ to A^θ defined by (7).

Theorem 4. For $g \in G$, $V(g^{-1})AV(g)$ is mapped by $\theta' = g\theta$.

Proof. $V(g^{-1})AV(g) =$

$$\int \theta |g^{-1}\theta\rangle\langle g^{-1}\theta| d\mu(\theta) = \int g\theta |\theta\rangle\langle \theta| d\mu(g\theta).$$

Use the invariance of μ . \square

Further properties of the mapping from θ to A may be developed in a similar way. The mapping corresponds to the usual way that the operators are allocated to observables in the quantum mechanical literature. But note that this mapping comes naturally here from the notions of conceptual variable and e-variables on which group actions are defined.

4.4 The spectral theorem and operators for functions of θ

Assume that θ is real-valued. Then based on the spectral theorem (e.g., [14]) we have that there exists a projectionvalued measure, E on Ω_Θ such that for $|v\rangle \in D(A)$

$$\langle v|A|v\rangle = \int_{\sigma(A)} \theta d\langle v|E(\theta)|v\rangle. \quad (8)$$

Here $\sigma(A)$ is the spectrum of A as defined in [14]. The case with a discrete spectrum is discussed in the next subsection.

A more informal way to write (8) is

$$A = \int_{\sigma(A)} \theta dE(\theta).$$

This defines a resolution of the identity

$$\int_{\sigma(A)} dE(\theta) = I. \quad (9)$$

From this, we can define the operator of an arbitrary Borel-measurable function of θ by

$$A^{f(\theta)} = \int f(\theta) dE(\theta). \quad (10)$$

Important special cases include $f(\theta) = I(\theta \in B)$ for sets B . Another important observation is the following: Any accessible variable can be written as $f(\theta)$, where θ is maximally accessible.

A further important case is connected to statistical inference theory in the way it is advocated in [1]. Assume that there are data z and a statistical model for these data of the form $P(X \in C|\theta)$ for sets C . Then a positive operator-valued measure (POVM) on the data space can be defined by

$$M(C) = \int P(X \in C|\theta) dE(\theta). \quad (11)$$

The density of M at a point z is called the likelihood effect in [1], and is the basis for the focused likelihood principle formulated there; see also Section 8 below.

Finally, given a probability measure with density $\pi(\theta)$ over the values of θ , one can define a density operator σ by

$$\sigma = \int \pi(\theta) dE(\theta). \quad (12)$$

In [1] the probability measure π was assumed to have one out of three possible interpretations: 1) as a Bayesian prior, 2) as a Bayesian posterior or 3) as a frequentist confidence distribution (see [13]).

4.5 The case of a purely discrete spectrum

Note that the construction in the previous subsections can be made for any accessible conceptual variable θ . Assume now that A has a purely discrete spectrum. Let the eigenvalues be $\{u_j\}$ and let the corresponding eigenspaces be $\{V_j\}$. The vectors of these eigenspaces are defined as quantum states, and as in [1], each eigenspace V_j is associated with a question ‘What is the value of θ ?’ together with a definite answer ‘ $\theta = u_j$ ’. This assumes that the set of values of θ can be reduced to this set of eigenvalues, which I will justify as follows. .

Theorem 5. *Assume that the group G is transitive on Ω_Θ . Then if one of the values of θ is an eigenvalue of A , all the values of θ are eigenvalues of A , and G is a permutation group on these eigenvalues.*

Proof. For each j , let $|j\rangle$ be an eigenvector of A with eigenvalue u_j , and let $g \in G$. By Theorem 3 we have that the operator $V(g^{-1})AV(g)$ is mapped by $g\theta$. Fix $\theta_0 \in \Omega_\Theta$, and assume that $\theta_0 = u_j$ for some j . We need to show that $g\theta_0$ is another eigenvalue for A , which follows from the fact that $|V(h^{-1})AV(h) - \lambda I| = |A - \lambda I|$, so that these two determinants have the same zeros.

It is crucial for this whole discussion that the group G is transitive on Ω_Θ . If this is not the case, one can use model reduction. The following principle has been proved useful in statistical inference. (See discussion and examples in [1], and an important application in [16, 17].)

Principle *When a group G is defined on the parameter space, every model reduction should be to an orbit or to a set of orbits of this group.*

In a quantum setting every model should be reduced to a single orbit. *This can be seen as a general principle for quantization.*

We also have the following:

Theorem 6. *Assume that the set of measurable values of θ is restricted to the above eigenvalues. Then θ is maximally accessible if and only if each eigenspace V_j is one-dimensional.*

Proof. The assertion that there exists an eigenspace that is not one-dimensional, is equivalent with the following: Some eigenvalue u_j correspond to at least two orthogonal eigenvectors $|j\rangle$ and $|i\rangle$. Based on the spectral theorem, the operator A corresponding to θ can be written as $\sum_r u_r P_r$, where P_r is the projection upon the eigenspace V_r . Now define a new e-variable ψ whose operator B has the following properties: If $r \neq j$, the eigenvalues and eigenspaces of B are equal to those of A . If $r = j$, B has two different eigenvalues on the two one-dimensional spaces spanned by $|j\rangle$ and $|i\rangle$, respectively, otherwise its eventual eigenvalues are equal to u_j in the space V_j . Then $\theta = \theta(\psi)$, and $\psi \neq \theta$ is inaccessible if and only if θ is maximally accessible. This construction is impossible if and only if all eigenspaces are one-dimensional.

5 Coupling different foci together

I will continue to discuss the case where the operator corresponding to θ has a discrete spectrum, but the whole discussion can be generalized.

5.1 Maximally accessible variables

In this section consider a Hilbert space \mathcal{H} of finite dimension d . Again let ϕ be an inaccessible conceptual variable. For an index set \mathcal{A} , focus on λ^a for $a \in \mathcal{A}$, a set of maximally accessible variables. Note that each λ^a corresponds to a unique operator A^a , and that this operator has the spectral decomposition

$$A^a = \sum_j u_j^a |a; j\rangle \langle a; j|. \quad (13)$$

By maximality, only one λ^a can be measured on the system at a given time. This is a manifestation of Niels Bohr's complementarity.

The following is proven in [1] under certain general technical conditions, and also specifically in the case of spin/ angular momentum: given a vector $|v\rangle \in \mathcal{H}$, there is at most one pair (a, j) such that $|a; j\rangle = |v\rangle$. The main interpretation in [1] is motivated as follows: Suppose the existence of such a vector $|v\rangle$ with $|v\rangle = |a; j\rangle$ for some a and j . Then the fact that the state of the system is $|v\rangle$ means that one has focused on a question ("What is the value of λ^a ?") and obtained the definite answer ($\lambda^a = u_j^a$.) The question can be associated with the orthonormal basis $\{|a; j\rangle; j = 1, 2, \dots, d\}$.

After this we are left with the problem of determining conditions under which *all* vectors $|v\rangle \in \mathcal{H}$ can be interpreted as above. This will require a rich index set \mathcal{A} . This problem will not be considered further here.

The following simple observation should also be noted: Trivially, every vector $|v\rangle$ is the eigenvector of *some* operators. Assume that there is one such operator A that is physically meaningful, and for which $|v\rangle$ is also a non-degenerate eigenvector. Let λ be a physical variable associated with A . Then $|v\rangle$ can again be interpreted as a question ("What is the value of λ ?") along with a definite answer to this question.

5.2 Non maximally accessible variables

First go back to the maximal symmetrical epistemic setting. Let again $\lambda^a = \lambda^a(\phi)$ be as in the previous subsection. Let t^a be an arbitrary function on the range of λ^a , and let us focus on $\theta^a = t^a(\lambda^a)$ for each $a \in \mathcal{A}$.

Let the Hilbert space be as in the previous subsection, and suppose that it has an orthonormal basis that can be written in the form $|a; i\rangle$ for $i = 1, \dots, d$. Let $\{u_i^a\}$ be the values of λ^a , and let $\{s_j^a\}$ be the values of θ^a . Define $C_j^a = \{i : t^a(u_i^a) = s_j^a\}$, and let V_j^a be the space spanned by $\{|a; i\rangle : i \in C_j^a\}$. Let Π_j^a be the projection upon V_j^a .

Then we have the following interpretation of any $|a; i\rangle \in V_j^a$. (1) the question: ‘What is the value of θ^a ?’ has been posed, and (2) we have obtained the answer $\theta^a = s_j^a$. Note that in this case, several pairs (a, i) correspond to a given vector $|v\rangle$.

From the above construction we may also define the operator connected to the e-variable θ^a as

$$A^a = \sum_j s_j^a \Pi_j^a = \sum_i t^a(u_i^a) |a; i\rangle \langle a; i|. \quad (14)$$

Note that this gives all possible states and all possible values corresponding to the accessible e-variable θ^a . Unless the function t^a is one-to-one, the operator A^a has no longer distinct eigenvalues.

6 The spin case

Let ϕ be the total spin vector of a particle, and let $\theta^a = \|\phi\| \cos(\phi, a)$ be the component of ϕ in the direction a . If we have a coordinate system, the components θ^x, θ^y and θ^z are of special interest. As pointed out by Yoh Tanimoto, the following example shows that these components are not permissible if K is the rotation group for some fixed $\|\phi\|$:

Take $\phi_1 = (1, 0, 0)$, $\phi_2 = (0, 1, 0)$, and let k be the rotation in the xz -plane such that $k\phi_1 = (0, 0, 1)$ and $k\phi_2 = (0, 1, 0)$. Then the z -components of ϕ_1 and ϕ_2 are equal, but the z -components of $k\phi_1$ and $k\phi_2$ are different.

Consider now a Stern-Gerlach experiment with a beam of particles in the z direction. Write in general $\phi = (\phi^1, \theta^z)$ where ϕ^1 is the spin component in the xy -plane. Let K_0 be the group of rotations of ϕ^1 in this plane for fixed $\|\phi^1\|$.

Proposition 2. *Any component θ^a in the xy -plane is a permissible function of ϕ^1 under K_0 .*

Proof. It is sufficient to consider θ^x . Let $k_0 \in K_0$ and $\theta^x(\phi_1^1) = \theta^x(\phi_2^1)$. The group element k_0 rotates the x -axis to a new direction x_0 and ϕ_1^1 and ϕ_2^1 to new vectors $k_0\phi_1^1$ and $k_0\phi_2^1$. By the geometry, the angle (x, ϕ_1^1) must be equal to angle $(x_0, k_0\phi_1^1)$, and angle (x, ϕ_2^1) must be equal to angle $(x_0, k_0\phi_2^1)$. Because the cosines of the unrotated angles are equal by assumption, the cosines of the rotated angles must also be equal. \square

Unit vectors in a plane as quantum state vectors have been discussed extensively in [32]. These vectors can be seen as elements of a real 2-dimensional Hilbert space, and they are uniquely determined by the angle they form with the x -axis. Using this as a

point of departure, density matrices and quantum operators (matrices) were defined in [32]. The authors noted that all real 2-dimensional matrices were generated by the unit matrix and the two real Pauli spin matrices σ_1 and σ_3 .

Of special interest here is the integral quantization as it is defined in [32]. It can be seen as a quantization of functions $\theta(\cdot)$ defined on a set Φ , concretely as a mapping from the space of functions to the space of linear operators on the Hilbert space. This map should satisfy

(i) *Linearity*. The map should be linear.

(ii) *Unity*. The function $\theta = 1$ is mapped to the identity operator.

(iii) *Reality*. A real function is mapped to a self-adjoint operator.

(iv) *Covariance*. If Φ is acted on by a group K , then there exist a unitary representation V of the group such that $A_{g(k)\theta} = V(k^{-1})A_\theta V(k)$ with $g(k)\theta(\phi) = \theta(k\phi)$.

If first a resolution of the identity is found as in (6), then all the above requirements are fulfilled by A given by (7). This is specialized to quantization of a circle in [32]. Rotations on the circle form a subgroup of rotations $SO(3)$ on the sphere, and there is a well-known homomorphism from the group $SU(2)$ of unitary 2×2 matrices with unit determinant to $SO(3)$. Hence we may consider the extended group K to be the group $SU(2)$, and take as a point of departure the irreducible representations of this group.

Spin coherent states and their connection to the group $SU(2)$ are discussed in thoroughly in [12] and [33]. I will follow parts of [12] without going into details. It is crucial that any irreducible representation D is given by a nonnegative integer or half-integer r : $D(k) = D^r(k)$, $\dim D^r = 2r + 1$. In the representation space $\mathcal{H} = \mathcal{H}_r$ the canonical basis $|r; m\rangle$ exists, where m runs from $-r$ to r in unit steps. The infinitesimal operators $A^\pm = A^x \pm A^y$, $A^0 = A^z$ of the group representation V^r satisfy the commutation relations

$$[A^0, A^\pm] = \pm A^\pm, \quad [A^-, A^+] = -2A^0. \quad (15)$$

The operators A^x , A^y and A^z are related to infinitesimal rotations around the x -axis, y -axis and z -axis, respectively, and we take $\mathbf{A} = (A^x, A^y, A^z)$. The representation space vectors $|r; m\rangle$ are eigenvectors for the operators A^0 and $\mathbf{A}^2 = (A^x)^2 + (A^y)^2 + (A^z)^2$:

$$A^0|r; m\rangle = m|r; m\rangle, \quad \mathbf{A}^2|r; m\rangle = r(r+1)|r; m\rangle. \quad (16)$$

The operator $\exp[i\omega(\mathbf{n} \cdot \mathbf{A})]$, $\|\mathbf{n}\| = 1$, describes the rotation by the angle ω around the axis directed along \mathbf{n} . In Ref. 5 this was used to describe the coherent states $D(k)|\phi_0\rangle$ in various ways. The ket vector $|\phi_0\rangle$ may be taken as $|r; m\rangle$ for a fixed m ; the simplest choice is $m = -r$.

7 More on coupling foci in the discrete case

Standard quantum theory usually pays considerable attention to observables with a discrete, often finite number d of values. The purpose of this Section is to give a precise theory for this, in particular to show that any set of maximally accessible conceptual variables under weak conditions can be associated with operators in the simple Hilbert space \mathbb{C}^d , where \mathbb{C} is the complex field.

Let θ^a for some index $a \in \mathcal{A}$ be a set of maximally accessible variables, each taking a finite number of values d . As in the spin example, a natural group K defined on an hypothetical inaccessible variable ϕ such that $\theta^a = \theta^a(\phi)$; $a \in \mathcal{A}$, may be so large that the accessible variables θ^a are not permissible. Therefore, start by looking at pairs of variables θ^a and θ^b . Let ϕ^{ab} be an inaccessible subvariable of ϕ such that $\theta^i = \theta^i(\phi^{ab})$; $i = a, b$, and let K_{ab} be the group K specialized to the range Φ^{ab} of ϕ^{ab} . To begin with, make the following assumption:

Assumption 1. θ^a and θ^b are permissible with respect to K_{ab} for each $a, b \in \mathcal{A}$.

This induces groups G^a and G^b on the range of θ^a and θ^b , respectively. Because these ranges are assumed to be finite, the groups are subgroups of the permutation group.

Assumption 2. G^a is transitive on the range Θ^a of θ^a for each $a \in \mathcal{A}$ and the isotropy group of G^a is trivial.

One way to ensure this, is to let G^a be the cyclic group on Θ^a . If this is not the case, one can go to a subgroup of the original group. Any representation of a group induces representations on subgroups.

Now by (7) there is an operator A^a connected to the variable θ^a . Because θ^a is assumed to be maximally accessible, then by Theorem 6 the eigenspaces of A^a are one-dimensional. By Theorem 5 the possible values of θ^a are the eigenvalues u_j^a of A^a . By the spectral theorem (13) holds, and we have a new resolution of the identity

$$I = \sum_{j=1}^d |a; j\rangle \langle a; j|. \quad (17)$$

These operators work on a Hilbert space $\mathcal{H}^a = \text{span}(|a; 1\rangle, \dots, |a; d\rangle)$. I will argue that under weak conditions $\mathcal{H}^a = \mathcal{H}^b$ may be chosen for the selected pair a, b . In fact, this Hilbert space can be taken as \mathbb{C}^d .

Axiom 1. There is a group homomorphism from the underlying group K to the group $SO(3)$.

We first let θ^a and θ^b take some special values, by analogy to the spin case. Let $r = (d-1)/2$ so that $d = 2r+1$ is the dimension of the corresponding spin irreducible representation. At the outset we know only that θ^a and θ^b each take d different values; to begin with, let these values be $-r, -r+1, \dots, r-1, r$. Consider an xz -plane where θ^a is plotted along the x -axis and θ^b along the z -axis. Let ϕ^{ab} be the angle between the x -axis and a given vector in the plane from the origin, and let K_{ab} be the group of rotations of this vector. Then by Proposition 2, Assumption 1 will hold. The discretization of θ^a and θ^b may be taken as a result of spectral decomposition of the operators A^a and A^b , respectively.

As background for this discussion, let $D(k')$ be the unique irreducible representation of $SU(2)$ and of $SO(3)$ of dimension $d = 2r+1$. Seen as matrices, the operators

$D(k')$ may be taken to act on the d -dimensional vector space \mathbb{C}^d . These matrix representation is discussed in many books, for instance Ref. 10. It is stated on op. cit., page 129 that D is a single-valued representation of $SO(3)$ when r is an integer, a double-valued representation when r is half of an odd integer.

There is a homomorphism from K to the group $SO(3)$ of three-dimensional rotations, and the group K_{ab} of rotations in the plane may be considered as a subgroup of this group. Hence the original representation $V(k)$ induces a representation $W(k_0)$ on \mathbb{C}^d of the group K_{ab} on the range Φ^{ab} of ϕ^{ab} . The vector space \mathcal{H}^a is d -dimensional, and is isomorphic to \mathbb{C}^d .

Let first r be half of an odd integer. The variable θ^a , which may be taken as $\sqrt{r(r+1)}\cos(\phi^{ab})$, is a function of ϕ^{ab} , but if θ^a is maximally accessible, ϕ^{ab} is inaccessible: The sign of ϕ^{ab} can not be recovered from θ^a .

We now fix $\phi^0 \in \Phi^{ab}$ and $|\theta^0\rangle \in \mathbb{C}^d$, and define $|\theta^a(\phi^{ab})\rangle = W(k_0)|\theta^0\rangle$ for a k_0 such that $\theta^a(\phi^{ab}) = \theta^a(k_0\phi^0) = g^a\theta^a(\phi^0)$ for some g^a by permissibility. Note that this is possible because the group K_{ab} of plane rotations is transitive over the range Φ^{ab} of ϕ^{ab} .

A similar construction can be made for $\theta^b = \sqrt{r(r+1)}\sin(\phi^{ab})$. Fix $\phi^{0b} \in \Phi^{ab}$ and $|\theta^{0b}\rangle \in \mathbb{C}^d$, and define $|\theta^b(\phi^{ab})\rangle = W(k_{0b})|\theta^{0b}\rangle$ for k_{0b} such that $\theta^b(\phi^{ab}) = \theta^a(k_{0b}\phi^{0b}) = g^b\theta^b(\phi^{0b})$ for some g^b by permissibility. These vectors span \mathbb{C}^d as k_{0b} varies. Since G^a and G^b may be taken as similar subgroups of the permutation group of d values, there is a one-to-one correspondence between the two groups. Under this correspondence we may take $k_{0b} = k_0$.

This construction will also work when D is a double-valued representation of $SO(3)$, i. e., when r is half of an odd integer. We can make the same construction when r is an integer. However, in this case special attention must be given to the values $\theta^a = \theta^b = 0$, when the angle ϕ^{ab} is undefined. This must be taken as a special value s , and K^{ab} must be extended to a transitive group on $\Phi^{ab} \cup \{s\}$. Note that the representation W need not be, and will not be, irreducible.

Lemma 3. *The representation $W(k_0)$ can chosen to be the same for a and for b when k_0 varies over K_{ab} .*

Proof. To start with, W can be considered a subrepresentation of the d -dimensional irreducible representation D of $SO(3)$ on \mathbb{C}^d . In the case where r is an integer, W must be extended, but this estension is the same for a and b . \square

Because all pairs can be selected, we have then proved

Proposition 3. *The Hilbert space $\mathcal{H}^a = \mathcal{H}^b$ can be taken to be the same for each $a \in \mathcal{A}$, and may be taken as \mathbb{C}^d . Each operator A^b acts on this Hilbert space.*

Any variable ψ taking d different values can be written in the form $\psi^a = \eta(\theta^a)$ for some one-to-one function η , where θ^a is as above, and its operator will be of the form

$$\sum_{j=1}^d \eta(u_j^a) |a; j\rangle \langle a; j|. \quad (18)$$

Its eigenvectors again span the same Hilbert space $\mathcal{H} = \mathbb{C}^d$. A change of notation then gives

Theorem 7. *Let $\{\theta^a(\phi); a \in \mathcal{A}\}$ be a set of maximally accessible variables, each taking d different values. Assume that the underlying group K on the range Φ of ϕ satisfies Axiom 1. Then each θ^a can be associated with an operator A^a on the Hilbert space \mathbb{C}^d such that the eigenvalues of A^a are all possible values of θ^a . The operators can be defined as in (7), where μ is the counting measure, and have spectral decompositions as (13).*

Since θ^a and θ^b are maximally accessible for $a \neq b$, sharp knowledge of θ^a rules out any knowledge of θ^b . The observer is in a ‘do not know’ situation on this variable.

Accessible variables that are not maximal, can be taken as functions of maximally accessible variables, and thus operators may be defined as in (18), see [1] for a discussion.

The question is when Axiom 1 can be taken to hold.

Corollary 1. *When \mathcal{A} takes 2 values, Theorem 7 will hold without assuming Axiom 1.*

Proof. This case was considered in the proof above. \square

Theorem 2 for this case of 2 values can be extended to the (position, momentum) case by first considering discretized variables, and then letting d tend to infinity. The relevant limiting operation is discussed in Section 5.3 of [1].

Corollary 2. *Consider any system of n spin vectors. Then Axiom 1 will hold.*

Proof. For $n=1$ there is a homomorphism from $SU(2)$ to $SO(3)$, which is equivalent to the statement that the spin vector is permissible with respect to $K = SU(2)$. For $n > 1$ use a general property of the permissibility concept: A vector function is permissible if and only if each component is permissible. \square

Theorem 2 for this case should be compared to parts of the results in the important paper [35].

Having thus obtained the Hilbert space formalism, we need assumptions which lead to the Born rule, and those that lead to the Schrödinger equation (see [1] for an outline of this). This will essentially complete the program of deriving quantum theory from reasonable assumptions.

8 Focused likelihood

Assumed that we have focused on some index $a \in \mathcal{A}$ and on an experiment with parameter θ^a and data x^a , both assumed to be discrete. Then the statistical model gives the likelihood $f(x^a|\theta^a)$, and we can define the likelihood effect

$$E^a(\mathbf{u}^a; x^a) = \sum_j f(x^a|\theta^a = u_j^a)\Pi_j^a, \quad (19)$$

where Π_j^a is the projection upon the eigenvector space of the operator A^a corresponding to the eigenvalue $\theta^a = u_j^a$. This is in agreement with an ordinary definition in quantum theory, where any operator $E = \sum_j p_j \Pi_j$ so that $0 \leq p_i \leq 1$ and $I = \sum_j \Pi_j$ is a resolution of the identity, is called an effect.

An observation x^a is called regular if $f(x^a | \theta^a = u_i^a) = f(x^a | \theta^a = u_j^a)$ implies $u_i^a = u_j^a$.

In [1] the following principle is proved from the ordinary likelihood principle:

The focused likelihood principle. *Consider two potential experiments a and b in some setting with equivalent contexts, and assume that the inaccessible variable ϕ is the same in both experiments. Assume that both observations are regular. Then the two observations x^a and x^b have equal likelihood effects in the two experiments if and only if*

A) *The set of projectors $\{\Pi_i^a\}$ is equal to the set of projectors $\{\Pi_j^b\}$, so that the resolution of the identity and hence the questions to nature implied by the two experiments are equal.*

B) *The two observations produce equivalent experimental evidence on the relevant parameter.*

9 The Born formula

As in [1] I start with an assumption of rationality, coupled to an experimentalist A . I do not see A as being perfectly rational, but during his experiment he adheres to certain ideals, and these ideals can be modeled by a superior actor D which is perfectly rational. We assume that D has priors found in some way, so that he can do a Bayesian analysis. The experimental evidence connected to D is then given by his posterior probabilities $q = q^a$ depending on the data x^a . By the focused likelihood principle, q must be a function of the likelihood effect $q = q(E^a)$, assuming the the observation x^a is regular.

Note that I do not assume that A has a prior and is able to do a Bayesian analysis, only that the ideal actor D is able to act as a Bayesian.

The perfect rationality of D is formulated in terms of hypothetical betting situations and

The Dutch Book Principle. *No choice of payoffs in a series of bets shall lead to a sure loss for the bettor.*

The setting described above is called a rational epistemic setting. The following theorem is proved in [1]:

Theorem 8. *Assume a rational epistemic setting. Let E_1, E_2, \dots be likelihood effects in this setting, and assume that $E_1 + E_2 + \dots$ also is an effect. Then*

$$q(E_1 + E_2 + \dots) = q(E_1) + q(E_2) + \dots \quad (20)$$

When $E_1, E_2; \dots$ belong to the same experiment, this is quite obvious, but it is proved also for the case where the experiments are different. A function $q(\cdot)$ satisfying the

countable additivity condition (20) and in addition $0 \leq q(E) \leq 1$ for all E and $q(I) = 1$, is called a generalized probability measure. I can now make use of the following theorem by Busch [41]:

Theorem 9. *Any generalized probability measure is of the form $q(E) = \text{trace}(\sigma E)$ for some density operator σ .*

The arguments above are valid for any sets of experiments in a rational epistemic setting, where $q = q(E)$ is some probability, a function of the likelihood effect, which in some way expresses the experimental evidence of the experiment connected to the likelihood effect E .

Define now a perfect experiment as one where the measurement uncertainty can be disregarded. The quantum mechanical literature operates very much with perfect experiments which give well-defined states $|j\rangle$. From the point of view of statistics, if, say the 99% confidence or credibility region of θ is the single point u_j , we can infer approximately that a perfect experiment has given the result $\theta = u_j$.

For perfect experiments the observations x are essentially equal the the relevant parameter values $\theta = u_j$, so we can without loss of generality assume that these observations are regular.

Assume now that we have prepared the system by a perfect experiment giving the result $\theta^a = u_i$ for some accessible variable θ^a . Let us then do a new perfect experiment, asking the question ‘What is the value of θ^b ?’ for a new accessible variable θ^b . Assume to begin with that both the accessible variables are maximal, so that the relevant eigenvalues are non-degenerate with eigenvectors $|a; i\rangle$, respectively $|b; j\rangle$. Then the following is proved in [1] from Theorem 9:

Theorem 10. Born’s formula. *Assume a rational epistemic setting. In the above situation we have*

$$P(\theta^b = u_j^b | \theta^a = u_i^a) = |\langle a; i | b; j \rangle|^2. \quad (21)$$

There are several generalizations of this Born formula. One can let the last eigenvalue be degenerate with projector Π_j^b on the eigenvector space corresponding to $\theta^b = u_j^b$. Then the righthand side of (21) should be replaced by $\langle a; j | \Pi_j^b | a; j \rangle$. And the first preparatorial experiment can be one where we only have probabilities for the different values u_j^a , resulting in a density operator σ^a . Then $P(\theta^b = u_j^b | \sigma^a) = \text{trace}(\Pi_j^b \sigma^a)$.

It is of some interest that Theorem 9, which is the basis, also is valid for non-perfect experiments. Following the same arguments, and assuming regularity, this leads to (36) and (37) below, which can be related to quantum measurement theory.

As an application of Born’s formula, we give the transition probabilities for spin 1/2 particles. I will, for a given direction a define $\theta^a = +1$ if the measured spin component by a perfect measurement for some given particle is $+\hbar/2$ in this direction, $\theta^a = -1$ if it is $-\hbar/2$. Assume that a and b are two directions in which the spin component can be measured.

Proposition 4. *For a spin 1/2 particle we have*

$$P(\theta^b = \pm 1 | \theta^a = +1) = \frac{1}{2}(1 \pm \cos(a \cdot b)). \quad (22)$$

This is proved in many textbooks, and also in [2].

The derivations of Sections 3-9 now give a starting point for the development of quantum theory as it is given in many textbooks and also in Chapter 5 of [1]. The Schrödinger equation is also developed from an epistemic point of view in [1].

10 Some statistical inference theory

Statistical inference theory was mentioned on several occasions above. I now give an introduction to 3 approaches to statistical inference. After this, I will discuss an example and some brief elements of quantum measurement theory, in order to illustrate more closely the connections between quantum theory and statistical inference theory, as I see it.

10.1 Frequentist and Bayesian inference

The basic notions of statistical inference are data X , varying in a space Ω_X , a vector or scalar parameter θ varying in a space Ω_Θ and a statistical model defined as follows: There is a σ -algebra \mathcal{E}_X of subsets of Ω_X , and for each $\theta \in \Omega_\Theta$ there is a probability measure P_X^θ on the measurable space $(\Omega_X, \mathcal{E}_X)$. Usually one assumes that this family of probability models is dominated: There exist a measure π on $(\Omega_X, \mathcal{E}_X)$ and a likelihood function $f(x|\theta)$ such that $dP_X^\theta = f(x|\theta)d\pi$.

In statistics, θ is thought to model the unknown feature that one is interested in, the state of the system in question. The data X are the potential observations, and the purpose of the modeling is multifold: To give a rough description of the data generating process; provide parameters that can be estimated from data; allow focusing on certain parameters; give a language for asking questions about nature, and give a possibility to study deviations from the model and choosing new models.

Here I will concentrate on asking questions about nature, in particular on estimating θ from observations. Given observed data $X = x$, one makes a decision $\delta(x)$, where the function δ is thought to vary on some action space Ω_A . From this point of view, one can say that statistical inference is the science of finding an optimal δ in concrete problems.

The simplest concept is that of point estimation: The parameter θ is estimated by a function of the data: $\hat{\theta}(x)$. The properties of this estimation procedure is evaluated by looking at the before-experimental situation and using the statistical model: With the stochastic variable X inserted, $\hat{\theta}(X)$ is called an estimator. One good property might be that the estimator is unbiased: $E^\theta(\hat{\theta}(X)) = \theta$ for all θ , exactly or perhaps approximately. Another good property may be that it has a small variance. These two properties are sometimes combined in the requirement that the estimator should have a

mean square error that is as small as possible, where the mean square error is defined as

$$MSE(\hat{\theta}(X)) = E^\theta((\hat{\theta}(X) - \theta)^2) = \text{Var}^\theta(\hat{\theta}(X)) + (E^\theta(\hat{\theta}(X) - \theta))^2. \quad (23)$$

In a more general decision situation one may have defined a loss function $\gamma(\theta, \delta(x))$ that one may wish to minimize in some sense. The first step is then to define the risk

$$R(\delta, \theta) = E^\theta(\gamma(\theta, \delta(X))). \quad (24)$$

A basic problem is that the risk in almost all cases cannot be minimized uniformly for all θ . Which way one proceeds with this problem, depends on what school one wants to adhere to. One way to go on, is to limit oneself in the point estimation case to unbiased estimators; for this theory and related theories, see [23].

A related problem on the case of a real-valued θ is to find an interval from data in which one believes that θ belongs. Again the solution depends on what school one wants to adhere to. I first describe the solution of the frequentist school.

One is in a situation where the data are given, but still one wants to use the statistical model as if one was in a pre-experimental situation. This presupposes an hypothetical future experiment completely equivalent to the experiment that one has performed, and that one evaluates probabilities connected to this future experiment. Given a confidence coefficient $1 - \alpha$, say 0.95, the aim is then to find lower and upper estimators $\underline{\theta}$ and $\bar{\theta}$ such that

$$P^\theta(\underline{\theta}(X) \leq \theta \leq \bar{\theta}(X)) = 1 - \alpha. \quad (25)$$

One then reports a confidence interval $[\underline{\theta}(x), \bar{\theta}(x)]$.

Again a related problem is that of hypothesis testing, say that one wants to test a null hypothesis $\theta = \theta_0$. An important concept is then that of the alternative hypothesis. Say that one considers a one-sided alternative $\theta > \theta_0$, has found a point estimator $\hat{\theta}$ for θ and the value $\hat{\theta}(x)$ for the experimental data at hand. One may then report a p -value

$$p = P^{\theta_0}(\hat{\theta}(X) > \hat{\theta}(x)), \quad (26)$$

referring to a hypothetical experiment under the null hypothesis, giving the result X . If the p -value is small, a traditional limit has been 0.05, the finding of the experiment is reported as being significant.

The p -values have been very much used in empirical research, and this use has been heavily attacked recently. One problem is that when many questions are addressed, in one or in several studies, the overall probability of getting a 'significant' result is large. The conclusion is that one should be very careful with the automatic use of p -values. Often the report of a confidence interval will be more informative.

In concrete experiments one often has data X that consists of several independent observations, in general one may have have very much data. One mechanism by which one can reduce data, is related to the concept of sufficiency. A function T of the data is called a statistic. A statistic T is called sufficient if the distribution of X , given T and θ is independent of θ . Under weak conditions one then has factorization of the likelihood

$$f(x|\theta) = h(x)g(T(x)|\theta). \quad (27)$$

Thus the essential part of the observation is $T(x)$; apart from this, the data from the model could have been obtained by independent computer simulation. This is related to the sufficiency principle: If data x_1 and x_2 have $T(x_1) = T(x_2)$, then x_1 and x_2 contain the same experimental evidence about θ . Here the concept ‘experimental evidence’ is left undefined.

Birnbaum showed (see [1]) that the sufficiency principle together with an equally intuitive conditionality principle imply the likelihood principle: If x_1 and x_2 have proportional likelihood functions, where the constant of proportionality is independent of θ , then x_1 and x_2 contain the same experimental evidence about θ . Thus all information about θ is contained in the likelihood function.

There have been discussions also around the likelihood principle. In [1] the view is advocated that the likelihood principle should be taken as conditional, given the context of the experiment.

A completely different approach to statistical inference is given by the Bayesian school. Bayesian inference is possible if one has a prior distribution $P(d\theta)$ of the parameter. Then one can define a Bayesian risk

$$B(\delta) = \int R(\delta, \theta)P(d\theta), \quad (28)$$

where $R(\delta, \theta)$ is given by (24), and the problem is simply to find the decision rule that minimizes $B(\delta)$. In the case of point estimation and quadratic loss function for given data, this leads to

$$\hat{\theta}(x) = \int \theta P(d\theta|x), \quad (29)$$

where

$$P(d\theta|x) = \frac{l(x|\theta)P(d\theta)}{\int l(x|\eta)P(d\eta)}. \quad (30)$$

The distribution (30) is called the posterior distribution of θ , and is a fundamental tool of Bayesian inference. (30) is an instance of Bayes’ formula.

Under the posterior distribution, the parameter Θ is a random quantity, and (29) may be written $\hat{\theta}(x) = E(\Theta|x)$.

The Bayesian concept corresponding to a confidence interval is that of a credibility interval: Assume that one can find $\theta_*(x)$ and $\theta^*(x)$ such that

$$P(\theta_*(x) \leq \Theta \leq \theta^*(x)|X = x) = 1 - \alpha. \quad (31)$$

Then $[\theta_*(x), \theta^*(x)]$ is called a credibility interval for Θ with credibility coefficient $1 - \alpha$. Note that, in contrast to (25), the probabilities are now computed directly from the posterior distribution of Θ .

The main problem of the Bayesian approach to inference is to find a prior $P(d\theta)$. A subjective Bayesian will say that this always can be specified by the user; sometimes one here makes use of a hypothetical betting construction. Another school relies on an ‘objective’ prior; one such approach assumes that there is a group defined on the parameter space.

Sometimes, but not always, one starts with a group G^* acting on the sample space Ω_X , so that this induces a group G on the parameter space by assuming that the model

parameter is uniquely determined by the model: $g\theta$ is defined by $P_X^{g\theta} = P_{gX}^\theta$. This defines a group homomorphism, sometimes an isomorphism. Also, there is a group G^{**} acting on the action space Ω_A , and it is assumed that there is an homomorphism from G^* to G^{**} defined through an permissible inference rule: $g^{**}\delta(x) = \delta(g^*x)$. If the action is an estimation procedure, and the homomorphism from G^* to G is an isomorphism, it is natural to assume also that this last homomorphism is an isomorphism. Also, one usually assumes that the loss function satisfies $\gamma(g\theta, g^{**}a) = \gamma(\theta, a)$. In the isomorphism case, I will use the symbols G, g instead of G^*, g^* and G^{**}, g^{**} .

An estimator $\hat{\theta}(X)$ is called equivariant if it transforms under the group in the same way as the parameter θ , i.e., $g\hat{\theta}(X) = \hat{\theta}(g^*X)$. More generally, a decision rule is called invariant if $g^{**}\delta(x) = \delta(g^*x)$ for all x . The risk function (24) of an invariant decision rule is constant on orbits of θ , see [7]. In the transitive case this leads to a unique invariant decision rule, see also [2].

When a group G is defined on the parameter space, there are many arguments to the effect that the right invariant measure connected to G should be used as an objective prior for θ . (See [2], for instance.) An argument sometimes raised against this, is that it very often leads to improper priors. As recently shown in [6], however, this problem can be solved by a slight relaxation of Kolmogorov's axioms if the marginal law of X is σ -finite: There are events B_1, B_2, \dots with $\Omega_X = \cup_i B_i$ and $P(X \in B_i) < \infty$ for $i = 1, 2, \dots$

For comparing statistical inference theory with quantum theory, we need to consider the case where both the parameter and the data are discrete. Then, with enough data, the credibility interval or the confidence interval may shrink to a single point θ_0 , so that $P(\Theta = \theta_0 | X = x) = 1 - \alpha$, respectively $P(\hat{\theta}(X) = \theta_0) = 1 - \alpha$. When α is small enough, we may then identify the post-experimental parameter value with θ_0 , and thus disregard experimental error. This explains why there is no distinction between data and parameters in quantum mechanical textbooks.

A classical statistical text treating both Bayesian and frequentist inference, is Berger [7]. Otherwise, there are many textbooks on statistical inference, for instance Bickel and Doksum [8]. Ferguson [34] is a classical text using a decision theoretical approach. For more on the case where a group is defined on the parameter space, see [2] and [15].

10.2 Fiducial theory in the group case

The discussion of this Section is based on [10]. Fiducial inference as an alternative to Bayesian inference was first proposed by Fisher [22], has been in discredit for many years, but has been the subject to several investigations recently. I use some space to it here because it is a distinctly new approach to statistical inference, different from frequentist and Bayesian inference.

Consider a statistical model $(\Omega_X, \mathcal{E}_X, P_X^\theta | \theta \in \Omega_\Theta)$, and assume that the loss of an action $a \in \Omega_A$ is of the form $l = \gamma(\theta, a)$. Both Θ and X are defined on an underlying abstract conditional probability space (Ω, \mathcal{E}, P) . The model parameter Θ is σ -finite.

Assume again that there is a group G defined on the parameter space. It is an essential assumption that the group G is transitive on Ω_Θ , and that the isotropy group is trivial. Then we can fix $\theta_0 \in \Omega_\Theta$ and establish a one-to-one correspondence between g and θ by $\theta = g\theta_0$. Consider the case where the group actions on Ω_Θ, Ω_X and Ω_A are isomorphic.

Let now U be a random quantity such that $P(U \in B|\theta) = P(X \in B|\theta = \theta_0)$ holds identically for all B and θ . The group invariance of the statistical model implies $P_X^\theta = P_{g_\theta X}^{\theta_0}$, and since by the definition of U it follows that the law of $g_\theta X$, given $\Theta = \theta_0$ is equal to the law of $g_\theta U$, given $\Theta = \theta$, this also equals $P_{g_\theta U}^\theta$. It follows that

$$(X|\Theta = \theta) \sim (g_\theta U|\Theta = \theta), \quad (32)$$

where the notation $(W_1|\Theta = \theta) \sim (W_2|\Theta = \theta)$ in general means just $P_{W_1}^\theta = P_{W_2}^\theta$.

Now following [10] we define

Definition 3 (Fiducial model). *Let Θ be a σ -finite random parameter defined on a space Ω_Θ . A fiducial model (U, ζ) is given by a random variable U defined on a space Ω_U and a measurable function $\zeta : \Omega_U \otimes \Omega_\Theta \rightarrow \Omega_X$, where Ω_X is the sample space of a statistical model $\{P_X^\theta|\theta \in \Omega_\Theta\}$. This is the fiducial model for the given statistical model if*

$$(\zeta(U, \Theta)|\Theta = \theta) \sim (X|\Theta = \theta) \quad (33)$$

In particular it follows from this definition and the construction given in connection to (32) above, that (U, ζ) with

$$\zeta(u, \theta) = g_\theta u \quad (34)$$

is a fiducial model for the statistical model $\{P_X^\theta|\theta \in \Omega_\Theta\}$ defined there.

It is allowed above that P_U^θ does depend on θ . However in this article, as mostly in [10], it is assumed that the fiducial model is conventional in the sense that P_U^θ does not depend on θ .

A fiducial model $(U; \zeta)$ is called simple if the fiducial equation $\zeta(u, \theta) = x$ has a unique solution $\theta^x(u)$ when solved for θ for all (u, x) . In the simple and conventional case the fiducial distribution is defined as the distribution of $\Theta^x = \theta^x(U)$ conditional on $\Theta = \theta$.

In [10] the following important theorem was proved:

Theorem 11. *In the above case the risk of an equivariant rule is determined by a fiducial distribution.*

In the setting above an equivariant rule is determined by $g_\theta \delta(x) = \delta(g_\theta x)$, where g_θ again is a group element acting first on Ω_A , then on Ω_X . The risk is defined from the loss function $l = \gamma(\theta, a)$ by

$$\rho = E(\gamma(\Theta, \delta(X))|\Theta = \theta). \quad (35)$$

Several examples of application of this result are given in [10].

10.3 Connection between approaches to statistical inference

For an outsider it may be confusing that there are several schools of statistical inference. Then it is reassuring that there are connections between these schools. First, in cases where there is very much data, the different theories give essentially the same result. In these cases, the effect of the choice of prior vanishes almost completely.

An even closer connection between the theories exists when there is a group G defined on the parameter space. Then there is a classical result by Fraser [24,25] and generalized by Taraldsen and Lindqvist [26] saying that the fiducial coincides with the posterior from the right invariant measure as a Bayesian prior.

There is also a close connection between Bayesian and frequentist inference in this case: In [2] the following general result is proved (recall the definition of a proper group in Section 2):

Proposition 1. *Assume that G is proper and is acting transitively both on Ω_X and on Ω_Θ . Let $\eta(\theta)$ be a one-dimensional continuous parametric function, and let $\hat{\eta}_1(x)$ and $\hat{\eta}_2(x)$ be two equivariant estimators of $\eta(\theta)$ such that $\hat{\eta}_1(x) < \hat{\eta}_2(x)$ for all x . Define $C(x) = \{\theta : \hat{\eta}_1(x) \leq \theta \leq \hat{\eta}_2(x)\}$. Then $C(x)$ is a credibility interval and a confidence interval of θ with the same credibility coefficient/ confidence coefficient,*

11 A macroscopic example

A very relevant question is now: Are all the results of Sections 3-9, including Born's formula, by necessity confined to the microworld? Recently, physicists have also become interested in larger systems where quantum mechanics is valid, see [36]. Of even more interest are the quantum models of cognition, see [37, 38]. As we have defined it, there is nothing microscopic about the epistemic setting. It may or may not be that the assumptions made above also are valid for some larger scale systems. The following example illustrates the point.

In a medical experiment, let μ_a, μ_b, μ_c and μ_d be continuous inaccessible parameters, the hypothetical effects of treatment a, b, c and d , respectively. Assume that the focus of the experiment is to compare treatment b with the mean effect of the other treatments, which is supposed to give the parameter $\frac{1}{3}(\mu_a + \mu_c + \mu_d)$. One wants to do a pairwise experiment, but it turns out that the maximal parameter which can be estimated, is

$$\theta^b = \text{sign}(\mu^b - \frac{1}{3}(\mu_a + \mu_c + \mu_d)).$$

(Imagine for example that one has four different ointments against rash. A patient is treated with ointment b on one side of his back; a mixture of the other ointments on the other side of his back. It is only possible to observe which side improves best, but this observation is assumed to be very accurate. One can in principle do the experiment on several patients, and select out the patients where the difference is clear.) This experiment is done on a selected set of experimental units, on whom it is known from earlier accurate experiments that the corresponding parameter

$$\theta^a = \text{sign}(\mu^a - \frac{1}{3}(\mu_b + \mu_c + \mu_d))$$

takes the value $+1$. In other words, one is interested in the probabilities

$$\pi = P(\theta^b = +1 | \theta^a = +1).$$

Consider first a Bayesian approach. Natural priors for μ_a, \dots, μ_d are independent normal distributions, $N(\nu, \sigma^2)$ with the same ν and σ . By location and scale invariance, there is no loss in generality by assuming $\nu = 0$ and $\sigma = 1$. Then the joint prior of $\zeta^a = \mu_a - \frac{1}{3}(\mu_b + \mu_c + \mu_d)$ and $\zeta^b = \mu_b - \frac{1}{3}(\mu_a + \mu_c + \mu_d)$ is multinormal with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} \frac{4}{3} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{3} \end{pmatrix}.$$

(A vector variable \mathbf{v} has a multinormal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if $\mathbf{a}'\mathbf{v}$ is $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for every constant vector \mathbf{a} .) A numerical calculation from this gives

$$\pi = P(\zeta^b > 0 | \zeta^a > 0) \approx 0.43.$$

This result can also be assumed to be valid when $\sigma \rightarrow \infty$, a case which in some sense can be considered as independent objective priors for μ_a, \dots, μ_d .

Now consider a rational epistemic setting for this experiment. Since again scale is irrelevant, a natural group on μ_a, \dots, μ_d is a 4-dimensional rotation group around a point (ν, \dots, ν) together with a translation of ν . Furthermore, ζ^a and ζ^b are contrasts, that is, linear combinations with coefficients adding to 0. The space of such contrasts is a 3-dimensional subspace of the original 4-dimensional space, and by a single orthogonal transformation, the relevant subset of the 4-dimensional rotations can be transformed into the group G of 3-dimensional rotations on this latter space, and the translation in ν is irrelevant. One such orthogonal transformation is given by

$$\begin{aligned} \psi_0 &= \frac{1}{2}(\mu_a + \mu_b + \mu_c + \mu_d), \\ \psi_1 &= \frac{1}{2}(-\mu_a - \mu_b + \mu_c + \mu_d), \\ \psi_2 &= \frac{1}{2}(-\mu_a + \mu_b - \mu_c + \mu_d), \\ \psi_3 &= \frac{1}{2}(-\mu_a + \mu_b + \mu_c - \mu_d). \end{aligned}$$

Let G be the group of rotations orthogonal to ψ_0 . We find

$$\begin{aligned} \zeta^a &= -\frac{2}{3}(\psi_1 + \psi_2 + \psi_3), \\ \zeta^b &= -\frac{2}{3}(\psi_1 - \psi_2 - \psi_3). \end{aligned}$$

The rotation group element transforming ζ^a into ζ^b is homomorphic under G to the rotation group element g_{ab} transforming $a = -\frac{1}{\sqrt{3}}(1, 1, 1)$ into $b = -\frac{1}{\sqrt{3}}(1, -1, -1)$. Let G^a be the maximal subgroup of G under which ζ^a is permissible. This is isomorphic with the group of rotations around a together with a reflection in the plane perpendicular to a , but the action on ζ^a is just a reflection. The orbits of this group are given by two-point sets $\{\pm c\}$. In conclusion, the whole situation is completely equivalent

to the spin-example of Proposition 4 and satisfies the assumptions of the symmetrical epistemic setting. Making the rationality assumption then implies:

$$\pi = P(\text{sign}(\zeta^b) = +1 | \text{sign}(\zeta^a) = +1) = \frac{1}{2}(1 + a \cdot b) = \frac{1}{3}.$$

I guess that many statisticians will prefer the Bayesian calculations here for the quantum theory calculations, which some may consider to have a more speculative foundation. But the prior chosen in this example must be considered somewhat arbitrary, and its 'objective' limit may lead to conceptual difficulties. Since experiments of this kind can in principle be done in practice - at least approximately, the question whether the Bayesian solution or the rational epistemic setting solution holds in such cases, must ultimately be seen as an empirical question.

The purpose of this example is not primarily to show how quantum theory applies in a macroscopic setting. Other, perhaps more interesting examples can be found in cognitive modeling [37, 38] and in economics (e.g., [39]). The purpose is more to indicate a border where quantum modeling may or may not be valid, i.e. where the assumptions made in this article may or may not hold.

12 Some quantum measurement theory

Consider first a simple measurement of a maximally accessible variable θ with associated operator $A = \sum_n u_n |n\rangle\langle n|$. Let the statistical model for the measurement be given by $P(X = x_j | \theta = u_n)$, assuming discrete data. Then assuming that the physical system first is prepared in some mixed state σ , one makes a measurement on θ , and the probability of obtaining the measurement result $X = x_j$ is according to a slight generalization of Born's formula

$$p_j = P(X = x_j | \sigma) = \text{trace}(M(j)\sigma), \quad (36)$$

where

$$M(j) = \sum_n P(X = x_j | \theta = u_n) |n\rangle\langle n|. \quad (37)$$

Then the final state must in some way be determined by $M(j)$, σ and p_j . The formula in question can be made consistent with the measurement theory of [18], given below, if one can find operators A_j such that $M(j) = A_j^\dagger A_j$.

This discontinuous change of state has caused much discussion in the literature, but with an epistemic interpretation of quantum theory it seems to be less problematic: It just represents an updating of the information that one has on θ .

According to [18], the most general measurement one can have on a system can be described as follows: Let $\{A_j\}$ be a set of operators satisfying $\sum_j A_j^\dagger A_j = I$. We want to measure a variable θ with possible values u_n . If σ is the state of the system before measurements, $\tilde{\sigma}_j$ is the state of the system after obtaining the measurement result $z = z_j$, and p_j is the probability of obtaining this result, then

$$\tilde{\sigma}_j = \frac{A_j \sigma A_j^\dagger}{p_j}, \quad (38)$$

$$p_j = \text{trace}(A_j^\dagger A_j \sigma). \quad (39)$$

In general, (38) and (39) give the inference rule implied by the quantum measurement. The arguments behind these formulae, as given in [18], is of interest. I will sketch a possible starting point of these arguments from a statistical perspective.

From this perspective, the measurement is described by the data X , and θ is the parameter of interest. During measurement one goes from the unknown variable (X, θ) to the known variable $(X = x, \hat{\theta}(x))$. This transition can be described by a group element $h \in G \otimes G$, and h can be represented by a unitary operator $U = U(h)$ acting on the vectors $|X = x_i\rangle|\theta = u_n\rangle$. In [18] similar arguments are given, using the concepts of probe and target. The existence of the unitary operator U can also be motivated by a classical theorem by Wigner [40].

The derivation in [18] then starts from an arbitrary value $X = x_0$ and an implied combined initial state given by the density matrix $\sigma_{comb} = |0\rangle\langle 0| \otimes \sigma$. The operator U is then written in terms of sub-blocks acting on the parameter space as $U = \sum_{j,j'} [j]\langle j'| \otimes A_{jj'}$, and it is argued that the first column of these blocks, $A_j = A_{j0}$ satisfy the resolution of the identity $I = \sum_j A_j^\dagger A_j$, and from this the final state (38) is argued for.

The special case when $\sigma = \sum P(n)|n\rangle\langle n|$ and all the operators A_j are diagonal in the vectors $|n\rangle$, i.e., $A_j = \sum_n A(j,n)|n\rangle\langle n|$, is also discussed in [18]. This case is shown to be equivalent with Bayesian inference with priors for θ equal to $P(n)$ and likelihood function $P(j|n) = |A(j,n)|^2$.

The point of this brief recapitulation of quantum measurement theory is to indicate that quantum measurement is a generalization of Bayesian inference, and that the prior in this inference theory could be anything, subjective or objective. The classical Bayesian inference corresponds to the case where everything can be formulated in terms of the projectors $|n\rangle\langle n|$, i.e., in a universe where we are only interested in a single resolution of the identity $I = \sum_n |n\rangle\langle n|$.

13 Discussion

The discussions given in Sections 3-9 here represent a completely new approach towards quantum theory, and these discussions should replace parts of Chapter 4 in my book [1]. The discussions are rather technical, but the foundation introduced based on conceptual inaccessible and accessible conceptual variables, seems to be much more intuitive than the ordinary formal Hilbert space foundation.

The resolutions of the identity (6,9,17 etc.) are crucial to this approach. They can be said to represent questions to nature: ‘What is the value of θ ?’ for an associated maximally accessible variable θ , and at the same time they give positive-operator valued measures. The resolution (6) is special, since the operators involved are not orthogonal projections.¹ In all other cases discussed here, the resolutions of the identity give directly projection-valued measures. Selecting one concrete projector Π (eigenspace V for a corresponding operator) gives a sharp answer to the relevant question to nature. The probabilities associated with these answers are given by the Born

¹By Neumark’s dilation theorem, the POVM here can be ‘lifted’ to a projection-valued measure.

rule $p = \text{trace}(\Pi\sigma)$ if we start with a state defined by the density operator σ . Different resolutions of the identity correspond to complementary questions.

The way I use ‘question’ here, a question may be composed of several elementary questions. This corresponds to the case where θ is a vector.

As mentioned in the previous Section, ordinary Bayesian inference corresponds to the case where we only have a single resolution of the identity. This is true in the discrete case, but it can be generalized. As thoroughly discussed in Section 10, there are also other approaches to statistical inference than the Bayesian one. It is interesting that both frequentist inference and fiducial inference are connected to ‘objective’ Bayesian inference, the one assuming a transitive group acting on the parameter space and the use of a right invariant prior.

Group theory, the assumption of a transitive group acting on the variable (parameter) space, is also important for the approach towards quantum theory advocated here. If we have a group defined on this space which is not transitive, model reduction - reduction to an orbit of the group - is called for. This is the approach to quantization advocated here. It is interesting that a similar model reduction also is useful in ordinary statistical inference.

The example in Section 11 shows that a quantum mechanical approach and a Bayesian statistical approach in general may give different solutions to an inference problem. It is suggested in this example that the question of which solution to choose, may ultimately be an empirical question.

In [37,38] it is proposed to use quantum theory in cognitive modeling and in decision theory. This use of quantum theory is important, but it is not discussed in the present paper. A parallel discussion may be provided, however. Conceptual variables are used in making decisions. These variables may be inaccessible if they are so comprehensive that no clear decision can be made. Then it may be a solution to focus on simpler, accessible decision variables.

The relationship between my approach and QBism was briefly discussed in Section 3. I also regard the Bayesian way of thinking as useful, in particular when we talk about an ideal observer. But I look upon Bayesianism as much wider than subjective Bayesianism; as mentioned above, ‘objective’ Bayesianism based on a right invariant prior is useful, both in applications and when discussing connections between different approaches to inference.

A subjective Bayesian considers hypothetical betting constructions to be a basis for making decisions. I rather regard ‘decision’ as a more primitive concept. A decision can be based on desires, beliefs or knowledges in some combinations, but we very often make decisions without having neither the time nor the opportunity to think of a hypothetical betting situation.

When asking questions to nature, two kinds of decisions are important. First we must decide how to focus before asking the question, and next, after an answer is obtained, we must decide how to interpret the answer.

14 Concluding remarks

Group theory and quantum mechanics are intimately connected, as discussed in details in [14] for example. In this article it is shown that the familiar Hilbert space formulation can be derived mathematically from a simple basis of groups acting on conceptual variables. The consequences of this is further discussed in [1]. This discussion also provides a link to statistical inference, as indicated in this paper.

From the viewpoint of purely statistical inference the accessible variables θ discussed in this paper are parameters. In many such situations also, it is useful to have a group of actions G defined on the parameter space; see for instance the discussion in [15]. In the present paper, the quantization of quantum mechanics is derived from the following principle: all model reductions in some given model should be to an orbit or to a set of orbits of the group G .

It is of some interest that the same criterion can be used to derive the statistical model corresponding to the partial least squares algorithm in chemometrics [16], and also to motivate the more general recently proposed envelope model [17].

This paper focuses on group symmetry in spaces of conceptual variables, particularly those referred to as parameters/e-variables in [1]. When data are present, in statistical inference one often starts with a group G^* on the data space (see for instance [15]). Referring to the statistical model $P^\theta(X \in B)$ for Borel-sets B in the data-space, the relationship is given by $P^{g\theta}(X \in B) = P^\theta(X \in g^*B)$. This induces a group homomorphism from G^* to G . In some cases, this is an isomorphism, and the same symbol g is often used in the data space and the parameter space. However, this is *not* the case when the data space is discrete and the parameter space is continuous, as when the model is given by a binomial distribution or a Poisson distribution. This case is studied in detail in [19]. In these cases, group-theoretical methods were first used to construct the Poisson family and the binomial family, and the basic tool is coherent states for certain groups (the Weyl-Heisenberg group in the Poisson case and the group $SU(2)$ in the binomial case). Finally, inference is studied by reversing the roles of data and parameter. The result in both cases is equivalent to Bayesian inference with a uniform prior on the parameter, which may be a coincidence. Taking this and the present paper as a point of departure, there seems to be a possibility to provide new ideas to symmetry-based approaches to statistical inference. As an extension of [19], a large class of probability distributions are shown to have connections to coherent states in [20].

In the present paper, the first axioms of quantum theory are derived from reasonable assumptions. As briefly stated in [1], one can perhaps expect after this, that such a relatively simple basis for quantum theory may facilitate a further discussion regarding its relationship to relativity theory. One can regard physical variables as conceptual variables, inaccessible inside black holes. However, such considerations go far beyond the scopes of both [1] and the present paper.

Acknowledgements

I am grateful to Gunnar Taraldsen for discussions on fiducial inference and to Anders Rygh Swensen for proposing the reference [34]. Also, I want to thank my family for everything they have sacrificed while I have been seeking an objective truth in science.

References

- [1] Helland, I.S.: *Epistemic Processes. A Basis for Statistics and Quantum Theory.* Springer, Berlin.(2018)
- [2] Helland, I.S.: *Steps Towards a Unified Basis for Scientific Models and Methods.* World Scientific, Singapore.(2010).
- [3] Nachbin, L.: *The Haar Integral.* Van Nostrand, Princeton, NJ. (1965)
- [4] Hewitt, E. and Ross, K.A.: *Abstract Harmonic Analysis, II.* Springer-Verlag, Berlin. (1970).
- [5] Wijsman, R.A.: *Invariant Measures on Groups and Their Use in Statistics.* Lecture Notes - Monograph Series 14, Institute of Mathematical Statistics, Hayward, California. (1990).
- [6] Taraldsen, G. and Lindqvist, B.H.: Improper priors are not improper. *The American Statistician* 64 (2), 154-158.
- [7] Berger, J.O.: *Statistical Decision Theory and Bayesian Inference,* 2nd Ed.. Springer-Verlag, New York (1985).
- [8] Bickel, P.J. and Doksum, K.A.: *Mathematical Statistics. Basic Ideas and Selected Topics.* Prentice Hall, New Jersey. (2001).
- [9] Helland, I.S.: Symmetry in a space of conceptual variables. *J. Math. Phys.* 60 (5) 052101 (2019).
- [10] Taraldsen, G. and Lindqvist, B.H.: Fiducial theory and optimal inference. *Annals of Statistics* 41 (1), 323-341.
- [11] Helland, I.S.: Quantum theory based on accessible and permissible conceptual variables. Submitted.
- [12] Perelomov, A.: *Generalized Coherent States and Their Applications.* Springer-Verlag, Berlin. (1986).
- [13] Schweder, T. and Hjort, N.L.: *Confidence, Likelihood, Probability. Statistical Inference with Confidence Distributions.* Cambridge University Press. (2016).
- [14] Hall, B.C.: *Quantum Theory for Mathematicians.* Graduate Texts in Mathematics, 267, Springer, Berlin. (2013).
- [15] Helland, I.S.: Statistical inference under symmetry. *International Statistical Review* 72, 409-422. (2004).
- [16] Helland, I.S., Sæbø, S. and Tjelmeland, H.: Near optimal prediction from relevant components. *Scandinavian Journal of Statistics* 39, 695-713. (2012).
- [17] Helland, I.S., Sæbø, S., Almøy, T. and Rimal, R.: Model and estimators for partial least squares. *Journal of Chemometrics* 32: e3044. (2018).
- [18] Jacobs, K.: *Quantum Measurement Theory and Its Applications.* Cambridge University Press, Cambridge. (2014).

- [19] Heller, B. and Wang, M.: Group invariant inferred distributions via noncommutative probability. *Recent Developments in Nonparametric Inference and Probability*, IMS Lecture Notes - Monograph Series 50, 1-19. (2006).
- [20] Ali, S.T. Gazeau, J.-P. and Heller, B.: Coherent states and Bayesian duality. *J. Phys. A: Math. Theor.* 41, 365302. (2008).
- [21] Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer-Verlag, New York. (1985).
- [22] Fisher, R.A.: Inverse probability. *Math. Proc. Cambridge Philos. Soc.* 26, 528-535. (1930).
- [23] Lehmann, E.L. and Casella, G.: *Theory of point estimation*. Springer, New York (1998).
- [24] Fraser, D.A.S.: On fiducial inference. *Annals of Mathematical Statistics* 32 (3), 661-676.
- [25] Fraser, D.A.S.: The fiducial method and invariance. *Biometrika* 48 (3/4), 261-280.
- [26] Taraldsen, G. and Lindqvist B.H.: Fiducial and posterior sampling. *Communications in Statistics - Theory and Methods* 44, 3754-3767. (2015).
- [27] Fuchs, C.A.: QBism, the perimeter of quantum Bayesianism arXiv: 1003.5209 [quant-ph]. (2010).
- [28] von Baeyer, H.C.: *QBism: The future of quantum physics*. Harvard University Press, Harvard.(2016).
- [29] Susskind, L. and Friedman, A.: *Quantum Mechanics. The Theoretical Minimum*. Basic Books, New York. (2014).
- [30] Schlosshauer, M., Kofler, J. and Zeilinger, A.: A snapshot of fundamental attitudes towards quantum mechanics. *Studies in History and Philosophy of Modern Physics* 44, 222-238. (2013).
- [31] Norsen, T. and Nelson, S.: Yet another snapshot of fundamental attitudes toward quantum mechanic. arXiv: 1306.4646v2 [quant-ph]. (2013).
- [32] H. Bergeron, E. M. F. Curado, J.-P. Gazeau, and L. M. C. S. Rodrigues: A baby Majorana quantum formalism. arXiv: 1701.04026v3 [quant-ph]. (2017).
- [33] J.-P. Gazeau: *Coherent States in Quantum Physics*. Wiley-VCH, Weinberg (2009).
- [34] T.S. Ferguson: *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, Cambridge, MA. (1967).
- [35] P.A. Höhn and C.S.P. Wever: Quantum theory from questions. *Phys. Rev. A* 95, 012102 (2017).
- [36] V. Vedral: Living in a quantum world. *Scientific American* 304 (6), 20-25. (2011).
- [37] E.M. Pothos and J.R. Busemeyer: Can quantum probability provide a new direction for cognitive modeling? With discussion. *Behavioral and Brain Sciences* 36, 255-327. (2013).
- [38] J.R. Busemeyer and P.D. Buza: *Quantum Models for Cognition and Decision*. Cambridge University Press, Cambridge. (2012).
- [39] D. Aerts, S. Sozzo and J. Tapia. Identifying quantum structures in the Ellsberg paradox. *Intern. J. Theor. Phys.* 53, 3666-3682. (2014).
- [40] E.P. Wigner: Group theory and its application to the quantum mechanics of atomic spectra. Academic, New York. (1959).
- [41] P. Busch: Quantum states and generalized observables. A simple proof of Gleason's Theorem. *Phys. Rev. Lett.* 91 (12), 120403.(2003).