Quantum theory of incompatible observations

Z Hradil† and J Summhammer

Atominstitut der Österreichischen Universitäten, Stadionallee 2, A-1020 Wien, Austria

Abstract. The question of what information about a quantum state may be inferred from a sequence of measurements made on it, is addressed. The main result is that maximum-likelihood estimation gives an arguably natural optimum approach in quantum theory. It singles out a particular choice of basis, in which the state is expressed. Various incompatible and incomplete observations made on the ensemble are treated as a complete generalized measurement. Consequently, the maximum likelihood (relative entropy) is the best measure for relating experimental data with theoretical predictions of quantum theory.

Quantum theory describes events on the most fundamental level currently available. The synthesis of information from mutually incompatible quantum measurements plays a key role in testing the structure of the theory. The purpose of this paper is to show the unique relationship between quantum theory and the mathematical statistics used to obtain optimal information from incompatible observations. Quantum theory prefers the relative entropy (maximum-likelihood principle) as the proper measure for evaluation of the distance between measured data and probabilities defined by quantum theory. For an experimentalist working in quantum physics it means that data should be fitted to the theory preferably using the maximum-likelihood estimation.

For the sake of simplicity and brevity we assume a discrete spectrum of the observed variable. This corresponds to the case of sharp and precise quantum measurements. Note, however, that these ideal assumptions are not detrimental. The more realistic case of observables with a continuous spectrum and finite experimental precision can be incorporated into this framework by replacing the corresponding projectors by a probability-valued operator measure (POVM) [1, 2]. Our main result is independent of a particular implementation of the quantum measurement and works in the very general case as well. In the following, we shall use the Dirac notation.

Let us start the exposition by considering repeated precise detection. This case is treated in standard textbooks of quantum theory [3]. Any observation is represented by a Hermitian operator $\hat{A}$, whose spectrum determines the possible results of the measurement

$$\hat{A}|a\rangle = a|a\rangle. \tag{1}$$

Eigenstates are orthogonal $\langle a|a'\rangle = \delta_{aa'}$ and the corresponding projectors provide the completeness relation

$$\sum_a |a\rangle\langle a| = \hat{1}. \tag{2}$$

† Permanent address: Department of Optics, Palacký University, 17 listopadu 50, 772 07 Olomouc, Czech Republic.

Projectors predict the probability for detecting a particular value of the q-variable a represented by the operator $\hat{A}$ as $p_a = \langle a|\rho|a \rangle$, provided that the system has been prepared in a quantum state ρ. This mathematical picture corresponds to the experimental reality in the following sense: when the measurement represented by the operator $\hat{A}$ is repeated N times on identical copies of the system, a particular output a is collected N_a times. The relative frequencies $f_a = \frac{N_a}{N}$ will sample the true probability as $f_a \rightarrow p_a$ fluctuating around them. The exact values are reproduced only in the asymptotic limit $N \rightarrow \infty$. The experimentalist's knowledge may be expressed in the form of a diagonal density matrix

$$\hat{\rho}_{est} = \sum_a f_a |a\rangle\langle a| \tag{3}$$

provided that error bars of the order of $1/\sqrt{N}$ are associated with the sampled relative frequencies. This should be understood as a mere rewriting of the experimental data $\{N, N_a\}$. Similar knowledge may be obtained by observations, which can be parametrized by operators diagonal in the $|a\rangle$ basis, i.e. by operators commuting with the operator $\hat{A}$. However, the possible measurement of *non-commuting* operators yields new information, which cannot be derived from the measurement of $\hat{A}$.

More may be learned about the unknown state of the system provided that observables corresponding to non-commuting operators will be registered on several copies of the same state. Let us assume that operators $\hat{A}_j$, $j = 1, 2, \ldots$ will be measured by probing the system N times together. Now, one expects to gain more than just the knowledge of the diagonal elements of the density matrix in some *a priori given* basis. This sequential measurement of non-commuting observables should be distinguished from the similar problem of approximate simultaneous measurement of non-commuting observables†. As in the former case of a single Hermitian operator, the latter case of various non-commuting operators may be represented by a series of projectors $|y_i\rangle\langle y_i|$. This should be accompanied by relative frequencies f_i indicating how many times a particular output i has been registered, $\sum_i f_i = 1$. Various states need not be orthogonal $\langle y_i|y_j \rangle \neq \delta_{ij}$, in contrast to the previous case of a single Hermitian operator. However, this substantial difference has deep consequences. The result of the measurement cannot be meaningfully represented in the same manner as previously. For example, direct linking of probabilities with relative frequencies used in standard reconstructions [4, 5] $\rho_{ii} = f_i$, $\rho_{ii} = \langle y_i|\hat{\rho}|y_i \rangle$, may appear as inconsistent, since the system of linear equations is overdetermined, in general.

Let us assume the existence of a quantum measure $F(\rho_{ii}|f_i)$, which parametrizes the distance between measured data and probabilities predicted by quantum theory. We will search for the state(s) located in the closest neighbourhood of the data. A general state may be parametrized in its diagonal basis as

$$\hat{\rho} = \sum_i r_i |\varphi_i\rangle\langle\varphi_i|. \tag{4}$$

The equation for the extremal states may be found analogously to the treatment developed in [6]. In particular, the formal necessary condition for extremal solution reads

$$\frac{\delta F(\rho_{ii}|f_i)}{\delta\hat{\rho}} = 0. \tag{5}$$

---

† The approximate simultaneous measurement of non-commuting operators $[\hat{A}\hat{B}] \neq 0$ can always be represented by measurement of commuting operators $\hat{A}$, $\hat{B}$ defined on the extended Hilbert space $\mathcal{H} = H_s \otimes H_a$, where H_s and H_a are the space of the original system and the space of the auxiliary field (ancilla), respectively.

Since the density matrix is parametrized according to the relation (4) with the help of independent (orthogonal) states $|\varphi_k\rangle$, the variation may be done along these rays yielding the system of coupled equations

$$\frac{\delta F(\rho_{ii}|f_i)}{\delta\langle\varphi_k|} = 0$$

for any allowed k. Using the relation

$$\frac{\delta F(\rho_{ii}|f_i)}{\delta\langle\varphi_k|} = \sum_i \frac{\partial F(\rho_{ii}|f_i)}{\partial\rho_{ii}}|y_i\rangle\langle y_i|\varphi_k\rangle r_k$$

the system of equations may be rewritten as the equation for the density matrix

$$\sum_i \frac{\partial F}{\partial\rho_{ii}}|y_i\rangle\langle y_i|\hat{\rho} = \lambda\hat{\rho} \tag{6}$$

where λ is a Lagrange multiplier. The normalization condition $\mathrm{Tr}\,\hat{\rho} = 1$ sets its value to

$$\lambda = \sum_i \frac{\partial F}{\partial\rho_{ii}}\rho_{ii}.$$

Any composed function $G(F(\rho_{ii}|f_i))$ fulfils the same extremal equation (6) with the Lagrange multiplier rescaled as $\lambda\frac{\mathrm{d}G}{\mathrm{d}F}$. Without loss of generality it is therefore enough to consider the normalization condition $\lambda = 1$.

The extremal equation (6) has the form of a decomposition of the identity operator on the subspace, where the density matrix is defined by

$$\sum_i \frac{\partial F}{\partial\rho_{ii}}|y_i\rangle\langle y_i| = \hat{1}_\rho. \tag{7}$$

This resembles the definition of POVM characterizing a generalized measurement [1, 2]. To link the above extremalization with quantum theory, let us postulate the natural condition for the quantum expectation value

$$Tr\left(\frac{\partial F}{\partial\rho_{ii}}|y_i\rangle\langle y_i|\hat{\rho}\right) = f_i. \tag{8}$$

This assumption is reasonable: the sequential incompatible observations with results f_i are regarded as a new measurement scheme, namely the measurement of the quantum state. The relation (8) interprets the detected data as quantum expectation values.

The quantum measure F then fulfils the differential equation

$$\frac{\partial F}{\partial\rho_{ii}}\rho_{ii} = f_i \tag{9}$$

and singles out the solution in the form

$$F(\rho_{ii}|f_i) = \sum_i f_i \ln\rho_{ii}. \tag{10}$$

This is nothing other than the log likelihood or Kullback–Leibler relative information†. Formal requirements of quantum theory, namely the interpretation of the extremal equation as a POVM, result in the concept of maximum likelihood in mathematical statistics. The analogy between

---

† Notice the asymmetry between the arguments f and p in definition of Kullback–Leibler relative information $K(f/p) = \sum_i f_i \ln(f_i/p_i)$. In the paper of Frieden [7] is the term Kullback–Leibler norm used for opposite ordering of data and probabilities. The case discussed here is called generalized Burg principle.

the standard quantum measurement associated with a single Hermitian operator, and a series of sequential measurements associated with many non-commuting operators is apparent now. The former determines the diagonal elements in the basis of orthonormal eigenvectors, whereas the latter estimates not only the diagonal elements, but the diagonalizing basis itself. This is the difference between the measurement of the quantum observable $\hat{A}$ and the measurement of the quantum state. In this sense maximum-likelihood estimation may be considered as a new quantum measurement. The observed quantum state is given by the solution of the nonlinear operator equation

$$\hat{R}(\rho)\hat{\rho} = \hat{\rho}$$
$$\hat{R}(\rho) = \sum_i \frac{f_i}{\rho_{ii}}|y_i\rangle\langle y_i|. \tag{11}$$

This is the completeness relation of a POVM with measured outputs $\{f_i\}$. The equation of this type resembles the so-called expectation–maximization (EM) algorithm [8] in mathematical statistics. The solution can be find using such an iterative method provided that basis is transformed in each step by a suitably chosen unitary transformation [9]. Maximum likelihood has recently been used for the solution of several problems in quantum theory [10]. Special cases of the solution (11) have been discussed for the operational phase concepts [11], the diagonal elements of the density matrix [12] and the reconstruction of the spin-$\frac{1}{2}$ state [13]. A numerical technique for maximum-likelihood estimation has been used in [14].

The quantum interpretation offers a new viewpoint on maximum-likelihood estimation. This method is customarily considered as just one of many estimation methods, unfortunately it is one of the most complicated ones. It is often considered as a subjective method, since likelihood quantifies the degree of belief in a certain hypothesis. Any physicist, an experimentalist above all, would perhaps use as their first choice the least-squares method for fitting theory and data [5]. Let us evaluate this as an illustrative counterexample. In this case $F(\rho_{ii}|f_i) = \sum_i (\rho_{ii} - f_i)^2$ and the extremal equation reads

$$2\sum_i (\rho_{ii} - f_i)|y_i\rangle\langle y_i|\hat{\rho} = \lambda\hat{\rho}$$
$$\lambda = 2\sum_i (\rho_{ii} - f_i)\rho_{ii}. \tag{12}$$

Equation (12) may again be interpreted as a completeness relation for the POVM

$$\hat{E}_i = 2\frac{(\rho_{ii} - f_i)}{\lambda}|y_i\rangle\langle y_i|.$$

However, the expectation value is given by a rather complicated implicit function of the measured data, since

$$\text{Tr}(\hat{\rho}\hat{E}_i) = 2\frac{(\rho_{ii} - f_i)\rho_{ii}}{\lambda}. \tag{13}$$

This does not mean that the least-squares method is incorrect, it only means that such a fitting does not reveal the structure of the quantum measurement. In this sense the maximum likelihood method is unique and exceptional.

There are several fundamental consequences of this result. According to Fisher's theorem [15], maximum-likelihood estimation is unbiased and achieves the Cramér–Rao bound asymptotically for large $N \to \infty$. As demonstrated here, for any finite N maximum likelihood may be interpreted as a quantum measurement. When seen this way, bias and the noise above the Cramér–Rao bound seem to be unpleasant but natural properties of quantum

systems. Maximum likelihood may set new bounds on distinguishability currently related to the Fisher information [16]. Fisher information corresponds to the Riemannian distinguishability metrics and may be naturally interpreted as the distance in the Hilbert space. Besides this, fundamental equations of quantum theory such as the Schrödinger equation, Klein–Gordon equation, Pauli equation, etc and other physical laws may be derived from the principle of minimum Fisher information [17, 18]. This again stresses the role of information in the formulation of fundamental laws of Nature.

Maximum likelihood generalizes the notion of POVM in the following sense. Actual measurements may be (and usually always are) incomplete. The maximum-likelihood approach shows how to complete them. Indeed, the extremal equation (11) represents the completeness relation. Therefore, any incomplete measurements (represented here by $|y_i\rangle\langle y_i|$) are complete in the subspace, where the completeness relation $\hat{R} = 1$ holds. POVM and estimated quantum state are mutually connected in their dependence on the type of observations and on their results. In particular, it is not necessary to consider only the special scheme for quantum state observation such as, for example, the mutually complementary eigenbases [19].

Note also that the formulation given here is free of any assumptions concerning the cost function and prior information [1, 2], since the conditions for estimation are different. In standard formulations the question is: 'What measurement is optimal for the given state and the given resolution measure (cost function)?' The decision rule than depends on both the prior distribution and the cost function. The amount of Shannon information acquired by a quantum measurement has been addressed by Massar and Popescu in [20]. It is a remarkable fact that there exists a model-independent answer to the question, how much can be learned (in bits) about a quantum state. In the formulation considered in this contribution the question was: 'What (mixed) state fits the given data obtained by the given measurement in the optimal way?' No prior information about the quantum state has been used, since in our opinion, such knowledge is not consistent with a quantum formulation. Any *a priori* information about a quantum state must come from another quantum measurement.

Acknowledgments

References

[1] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[2] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[3] Sakurai J J 1994 *Modern Quantum Mechanics* (Reading, MA: Addison-Wesley)
[4] Leonhardt U 1997 *Measuring of the Quantum State of Light* (Cambridge: Cambridge University Press)
[5] Welsch D-G, Vogel W and Opatrný T Homodyne detection and quantum state reconstruction *Progress in Optics* vol 39, ed E Wolf
[6] Hradil Z 1997 *Phys. Rev.* A **55** R1561
    Hradil Z, Summhammer J and Rauch H 1999 *Phys. Lett.* A **261** 20
[7] Kullback S and Leibler R A 1951 *Ann. Math. Stat.* **22** 79
    Frieden B R 1985 *Maximum Entropy and Bayesian Methods in Inverse Problems* ed C R Smith and W T Grandy Jr (Dordrecht: Reidel) p 133
[8] Dempster A P, Laird N M and Rubin D B 1977 *J. R. Stat. Soc.* B **39** 1
    Vardi Y and Lee D 1993 *J. R. Stat. Soc.* B **55** 569
[9] Hradil Z, Řeháček J and Ježek M 2000 in preparation

[10] Shapiro J H, Shepard S R and Wong N C 1989 *Phys. Rev. Lett.* **62** 2377
      Lane A S, Braunstein S L and Caves C M 1993 *Phys. Rev.* A **47** 1667
[11] Řeháček J, Hradil Z, Zawisky M, Pascazio S, Rauch H and Peřina J 1999 *Phys. Rev.* A **60** 473
[12] Hradil Z and Myška R 1998 *Acta Phys. Slov.* **48** 199
[13] Hradil Z, Summhammer J, Badurek G and Rauch H 2000 *Phys. Rev.* A **62** 014101
[14] Banaszek K, Ariano G M D, Paris M G A and Sacchi M F 1999 *Phys. Rev.* A **61** 010304(R)
[15] Fisher R A 1925 *Proc. Camb. Soc.* **22** 700
[16] Wootters W K 1981 *Phys. Rev.* D **23** 357
      Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
[17] Reginatto M 1998 *Phys. Rev.* A **58** 1775
      Reginatto M 1998 *Phys. Lett.* A **249** 355
[18] Frieden B R 1999 *Physics from Fisher Information* (Cambridge: Cambridge University Press)
[19] Wootters W K 1986 *Found. Phys.* **16** 391
[20] Massar S and Popescu S 2000 *Phys. Rev.* A **61** 062303

Annotations from 114201dd.TEX typeset.pdf

Page 5

Annotation 1; Label: Author query; Date: 25/09/2000 16:13:58
Please provide the year of publication for ref 5

Annotation 2; Label: Author query; Date: 25/09/2000 16:14:16
Any further details for ref 9, in preparation?