Remarks on some basic issues in quantum mechanics

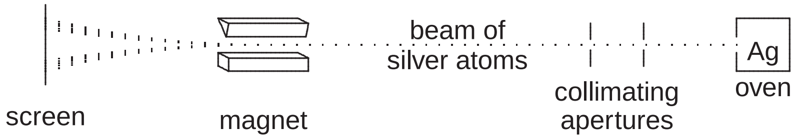
views from the past millennium

[Zeitschrift für Naturforschung **54a** (1999) 11–32]

Berthold-Georg Englert
Centre for Quantum Technologies (CQT) & Department of Physics
National University of Singapore

Events

Quantum mechanics can be used to predict probabilities — probabilities for what? For events!



Rudolf Haag, Commun. Math. Phys. **132** (1990) 245–257:

Events

– are localized in space and in time
– have an element of irreversibility
– are linked by particles, so that a causal history evolves
– are needed to give meaning to a space-time structure
– typically need to be amplified to be noticeable by human observers

Short-range interactions are crucial for the localization of events.

Nonlocality?

A referee report:

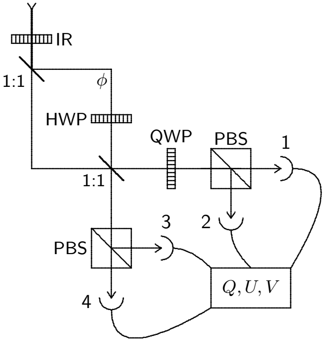
The authors claim that they have a proposal for an experiment that would demonstrate non-locality for a single photon. [. . .] there is no fundamental non-locality in quantum phenomena. Quantum mechanical processes are fundamentally local in the very definite sense that observations in space-like separated regions are independent. Technically speaking, the observables (elements of an operator algebra) of one region commute with all observables of the other region. Rudolf Haag's monograph on *Local Quantum Physics* (Springer, 1992) is recommended reading. All interactions considered in [. . .] are of the usual local kind.

The arbiter's comment:

The reviewer is correct to say that there is no fundamental non-locality in QM in the sense that information cannot be transmitted between space-like separated observers. However, in the present context, the phrase non-local is really a shorthand for "any realistic hidden variable theory capable of reproducing the results would have to be non-local". Workers in the field take this for granted, . . .

Generalized measurements

Single-photon polarimetry:



Probability of a click in the kth detector: $p_k = \text{tr}\{\Pi_k \rho\}$ where ρ is the statistical operator for the polarization of the incoming photon and Π_k is the probability operator for the kth detector.
The Π_ks make up a probability-operator measurement (POM), *vulgo* a positive-operator-valued measure (POVM):

$$\Pi_k \geq 0\,, \qquad \sum_k \Pi_k = 1$$

Murky Interpretation?

From the Introduction to a recent paper:

It is remarkable, that a century after the discovery of quantum mechanics, it seems that we are no closer to a consensus about its interpretation, than we were in the beginning. The collapse of the quantum state at the process of measurement which appears in all textbooks of quantum theory does not have an unambiguous definition and a reasonable explanation.
[. . .] some radical changes in our classical understanding of reality have to be made; e.g. constructing a physical process of collapse, accepting the existence of parallel worlds, or adding non-local hidden variables.

Rebuttal:

In fact, the is no lack of consensus, because the interpretation of a physical theory is simply the link between the mathematical symbols and the physical phenomena, such as $p_k = \mathrm{tr}\,\{\Pi_k \rho\}$ (Born's rule).
If you endow the symbols with more meaning than that, you yourself are responsible for the consequences — and don't blame quantum mechanics if you end up in dire straits.

Two kinds of evolution?

Folklore: “According to von Neumann there are two kind of evolution: unitary evolution between measurements, and sudden collapse at the time of measurement.”

Really? Reconsider $p = \text{tr}\,\{\Pi\rho\}$ where Π and ρ are functions of the dynamical variables $Z(t)$ with, possibly, a parametric time dependence as well, and all such operator-valued functions *evolve* in accordance with Heisenberg’s equation of motion,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(Z(t),t) = \frac{\partial}{\partial t} f(Z(t),t) + \frac{1}{\mathrm{i}\hbar}[f(Z(t),t), H(Z(t),t)]\,.$$

In particular, we have $\frac{\mathrm{d}}{\mathrm{d}t}\rho = 0$: the statistical operator is constant, as it should be because it represents our knowledge about the preparation of the system: $\rho(Z(t),t) = \rho(Z(t_0),t_0)$.

The “sudden collapse” is the state reduction by which we update ρ when we learn something new about the system.
State reduction is not a physical process, it is not *evolution*.

State reduction . . .

. . . is not particular to quantum mechanics. It is a book-keeping device of all statistical formalisms.

My wallet before:

$$	0	1	2	3	4	5	6	7	8	9	10
prob	0	0	0.5	0	0	0	0	0	0	0	0.5

My wallet after:

$$	0	1	2	3	4	5	6	7	8	9	10
prob	0	0	0	0	0	0	0	1	0	0	0

Action at a distance?

A recent news item:
Quantum theory makes the distinctive prediction that non-local correlations are instant: for example, a measurement of the polarization of one of a pair of quantum-entangled photons should immediately set the polarization of the other, no matter how far the photons are apart, without either photon's polarization being in any way predetermined.

Rebuttal: I told you: "you yourself are responsible for the consequences", I really did.

Here: Don't make the mistake of regarding the statistical operator (or the wave function for that matter) as a physical object. There is Alice's statistical operator for Bob's photon, and there is Bob's statistical operator for Bob's photon, and they can very well be different and both correct.

The Heisenberg cut

The cut is needed. By its nature, it cannot be precise.

Schrödinger's ~~cat~~ coin

Two macro-states: $\rho_1 = A^\dagger A\,, \quad \rho_2 = B^\dagger B$
Superpositions: $\rho_\pm = \frac{1}{2}(A \pm B)^\dagger (A \pm B) = \frac{1}{2}(\rho_1 + \rho_2) \pm \frac{1}{2}(A^\dagger B + B^\dagger A)$

No way to distinguish between ρ_+ and ρ_- because you wouldn't know how; in addition, the cross terms are ineffective (G. Süßmann 1958, A. Peres 1980, …); the phenomenology is that of $\rho = \frac{1}{2}(\rho_1 + \rho_2)$.

"But, in principle, I could … Dirac, von Neumann … one-to-one correspondence …"
No, you cannot.

For Heisenberg's dog, see J. Mod. Opt. **45** (1998) 701–711.

Quantum state estimation

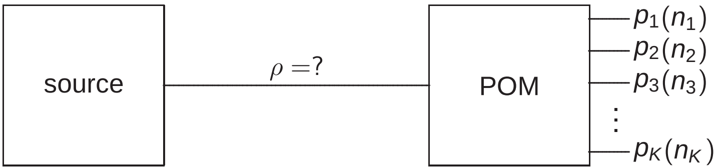
Recent work with

Ng Hui Khoon (postdoc, Singapore)

Zhu Hunangjun, Teo Yong Siah (PhD students, Singapore)
Benjamin Phuah (3rd year student, Singapore)

in collaboration with Z. Hradil, J. Řeháček, and B. Stoklasa (Olomouc) and D. Mogilevtsev (Minsk)

Scenario of quantum state estimation



The source emits quantum objects whose relevant degrees of freedom are described by the "true" statistical operator ρ, which is unknown.

The probability-operator measurement (POM) has K outcomes Π_k that give rise to the "true" detection probabilities $p_k = \text{tr}\{\rho\Pi_k\}$.

The actual data consist of $n_1, n_2, \ldots, n_K$ detector clicks in one particular sequence upon measuring a total of $N = n_1 + n_2 + \cdots + n_K$ copies.

State estimation: Exploit the data for an educated guess about ρ.

Principles of quantum state estimation

1– Be guided by common sense and the methods of classical statistical inference.*

2a– Estimate event probabilities from the data, after measuring N copies.

2b– Determine the estimator $\widehat{\rho}$ of the state from the estimated probabilities $\widehat{p}_1, \widehat{p}_2, \widehat{p}_3, \ldots$ and, if necessary, invoke additional criteria (such as Jaynes's maximum-entropy criterion).

Note: $\widehat{p}_k \to p_k^{(\mathrm{true})}$ for $N \to \infty$ ("consistency" — largely a tautology).

*Read Edwin Jaynes's *Probability Theory — The Logic of Science* and don't ignore his advice.

Tossing a coin (1)

Parameter to be estimated: p = probability of getting “head”.

Observed data: n times “head” and $N - n$ times “tail”.

Likelihood for the observed outcome: $\mathcal{L}(n, N-n|p) = p^n(1-p)^{N-n}$.

Maximum-likelihood (MaxLik) estimator:

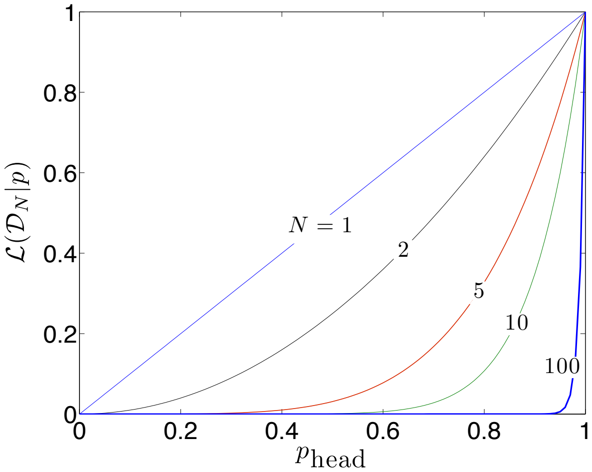
$$\widehat{p} \quad \text{such that} \quad \mathcal{L}(n, N-n|\widehat{p}) = \max_p \mathcal{L}(n, N-n|p).$$

One finds $\widehat{p}_{\mathrm{ML}} = \dfrac{n}{N}$ = relative frequency for “head”.

Note: Only the relative frequency matters.

Tossing a coin (2)

Likelihood for heads only and no tails



Tossing a K-sided die

Parameters to be estimated: probabilities $p_1, p_2 \dots, p_K$ with

$$\sum_{k=1}^{K} p_k = 1.$$

Observed data: $n_1, n_2, \dots, n_K$ occurrences for $N = \sum_{k=1}^{K} n_k$ tosses.

Likelihood for the observed outcome:

$$\mathcal{L}(n|p) \equiv \mathcal{L}(n_1, \dots, n_k|p_1, \dots, p_k) = \prod_{k=1}^{K} {p_k}^{n_k}.$$

MaxLik estimator: $\widehat{p}$ such that $\mathcal{L}(n|\widehat{p}) = \max_p \mathcal{L}(n|p)$.

One finds $\quad (p_k)_{\mathrm{ML}} = \frac{n_k}{N} =$ relative frequency for kth outcome.

Note: Only the relative frequencies matter.

Bayesian mean estimator

Idea ("applied common sense"): Regard the likelihood as a relative weight for the true values.

Estimator: $\widehat{p}_{\rm ME} = \dfrac{\int (\mathrm{d}p)\, \mathcal{L}(n|p)\, p}{\int (\mathrm{d}p)\, \mathcal{L}(n|p)}$ with $(\mathrm{d}p)$ = a-priori weight assigned to the parameter space element associated with $\mathrm{d}p_1\, \mathrm{d}p_2 \cdots \mathrm{d}p_K$.

A-priori weight: $(\mathrm{d}p) = \mathrm{d}p_1\, \mathrm{d}p_2 \cdots \mathrm{d}p_K\, f(p)$ with Bayesian prior $f(p)$.

Rules and regulations:

$f(p) > 0$ for all permissible p;

$f(p) = 0$ for all non-permissible p;

$\int (\mathrm{d}p)\, \mathcal{L}(n|p) < \infty$ when $n \neq 0$.

K-sided die revisited

Simplest prior: $f(p) = \delta\Big(1 - \sum_k p_k\Big)\big(p_1 p_2 \cdots p_K\big)^{\beta-1}$ with $\beta > 0$

(leaving conditions $p_k \geq 0$ implicit).

This gives $(\widehat{p}_k)_{\text{ME}}^{(\beta)} = \dfrac{n_k + \beta}{N + \beta K}$.

Note 1: We recover $(p_k)_{\text{ML}} = \dfrac{n_k}{N}$ in the limit $\beta \to 0$.

Note 2: The ME estimator looks *as if* we were computing the ML estimator for counts $n_k + \beta$, that is after adding β fake counts to each outcome. Such “add-β” estimators are sometimes used as ad-hoc corrections for ML estimators that are conceived as implausible.

Can we get non-ML estimators without any ad-hoc features?

Estimation error — Estimation risk

Error $=$ Distance between estimator and true state.

Here, squared Euclidean distance:

$$E(p,\widehat{p}) = \sum_{k=1}^{K} (p_k - \widehat{p}_k)^2$$

where $\widehat{p}$ is the estimator for data $n = \{n_1, n_2, \dots\}$ determined by some chosen strategy (ML, ME-β, ...).

Risk $=$ Average error

$$R_f(p) = \sum_n \frac{(n_1 + n_2 + \cdots + n_K)!}{n_1!\, n_2! \cdots n_K!} \mathcal{L}(n|p) E(p, \widehat{p}_f)$$

where $\widehat{p}_f$ is the estimator for prior $f(p)$.

Constant-risk estimation — Minimax estimation

Constant-risk estimation: Choose the prior $f(p)$ such that $R_f(p)$ is the same for all $p = \{p_1, p_2, \dots, p_K\}$.

Minimax estimation: Choose the prior $f(p)$ such that the largest risk is minimized:

$$\text{Optimal } f_0 \text{ such that } \quad \min_f \max_p R_f(p) = \max_p R_{f_0}(p).$$

Theorem: A constant-risk estimator, if it exists, is also a minimax estimator.

K-sided die revisited again

ML-β estimator: $(\widehat{p}_k)^{(\beta)}_{\text{ME}} = \dfrac{n_k + \beta}{N + \beta K}$.

Constant-risk estimator: $(\widehat{p}_k)_{\text{CR}} = (\widehat{p}_k)^{(\beta=\sqrt{N}/K)}_{\text{ME}} = \dfrac{n_k + \sqrt{N}/K}{N + \sqrt{N}\,K}$.

Minimax estimator: $(\widehat{p}_k)_{\text{MM}} = (\widehat{p}_k)_{\text{CR}}$.

Note 1: The optimal prior depends on N, the total number of tosses.

Note 2: The constant-risk/minimax estimator is

$$(\widehat{p}_k)_{\text{MM}} = (\widehat{p}_k)_{\text{CR}} = \frac{1}{K} a_N + \frac{n_k}{N} b_N$$

with $a_N = \dfrac{1}{\sqrt{N}+1}$ and $b_N = \dfrac{1}{1 + 1/\sqrt{N}}$.

Note 3: The MM estimator approaches the ML estimator for large N.

Note 4: The constant risk is $\propto \dfrac{1}{N}$ for large N.

Quantum example: Minimal qubit tomography (1)

Quantum state: $\rho = \frac{1}{2}(1 + \vec{s} \cdot \vec{\sigma})$.

Tetrahedron measurement: $\sum_{k=1}^{4} \Pi_k = 1$ with $\Pi_k = \frac{1}{4}(1 + \vec{t}_k \cdot \vec{\sigma})$

$$\text{where } \vec{t}_j \cdot \vec{t}_k = \begin{cases} 1 & \text{for } j = k, \\ -\frac{1}{3} & \text{for } j \neq k. \end{cases}$$

The probabilities $p_k = \operatorname{tr}\{\Pi_k \rho\} = \frac{1}{4}(1 + \vec{t}_j \cdot \vec{s})$ are constrained

$$\text{by } \sum_{k=1}^{4} p_k^2 \leq \frac{1}{3}.$$

The 4-sided die has no such constraint, but we note that

$$\sum_{k=1}^{4} (\widehat{p}_k)^2_{\mathrm{MM}} \leq 1 - \frac{3}{4}(1 - b_N)^2 < 1.$$

Quantum example: Minimal qubit tomography (2)

The constraint $\sum_{k=1}^{4} p_k^2 \leq \frac{1}{3}$ has to be obeyed by the prior $f(p)$, so that the priors that are useful for the 4-sided die are not applicable to the qubit.

Consequence 1: The simple $\widehat{p}_k = \frac{n_k}{N}$ may not be permissible, so that MaxLik estimation is more complicated. But it can be done rather easily, and small-N ML estimators tend to be implausible (we get rank-deficient states often).

Consequence 2: There is no corresponding constant-risk estimator for the qubit case. And the minimax problem, in its general form, does not have a known solution.

Quantum example: Minimal qubit tomography (3)

Likelihood function:
$\mathcal{L}(n|p) = p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}$, same as for the 4-sided die.

Simplest prior:
$f(p) = \delta\Big(1 - \sum_k p_k\Big)\big(p_1 p_2 p_3 p_4\big)^{\beta-1}$ for the 4-sided die.

Two analogs for the qubit:

(a) $f(p) = \delta\Big(1 - \sum_k p_k\Big)\eta\Big(1 - 3\sum_k p_k^2\Big)\big(p_1 p_2 p_3 p_4\big)^{\beta-1}$
is measurement specific.

(b) $f(p) = \delta\Big(1 - \sum_k p_k\Big)\eta\Big(1 - 3\sum_k p_k^2\Big)\big(\det\{\rho\}\big)^{\beta-1}$
is state specific.

There is no obvious preference for one of them, and the resulting difference in the minimax estimators is insignificant.

Qubit estimation: An act of desperation (1)

Estimator for the 4-sided die: $(\widehat{p}_k)_{\mathrm{CR}} = (\widehat{p}_k)_{\mathrm{MM}} = \frac{1}{K} a_N + \frac{n_k}{N} b_N$.

Ansatz for qubit (a bit of white noise admixed):

$$\widehat{p}_k = (1-\lambda)(\widehat{p}_k)_{\mathrm{MM}} + \frac{1}{4}\lambda$$

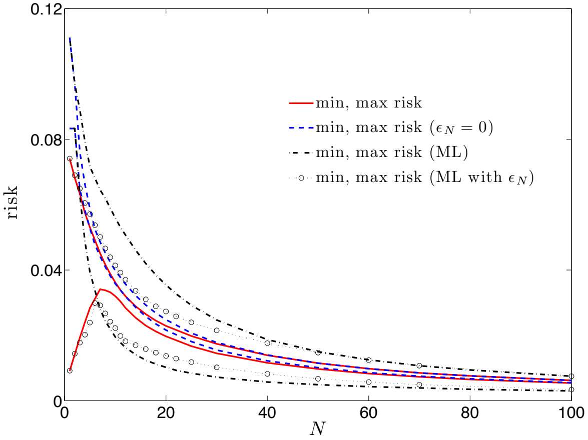
with $\lambda = 0$ for most n and $\lambda > 0$ for those n for which $\widehat{p}_{\mathrm{MM}}$ is unphysical.

Condition imposed: $\sum_{k=1}^{4} (\widehat{p}_k)^2 \leq \frac{1-\epsilon_N}{3}$ with pre-chosen $\epsilon_N > 0$ and smallest $\lambda \geq 0$ such that the condition is obeyed.

Minimax estimator: Determine the optimal value of ϵ_N such that we have minimax estimation within this family of estimators.

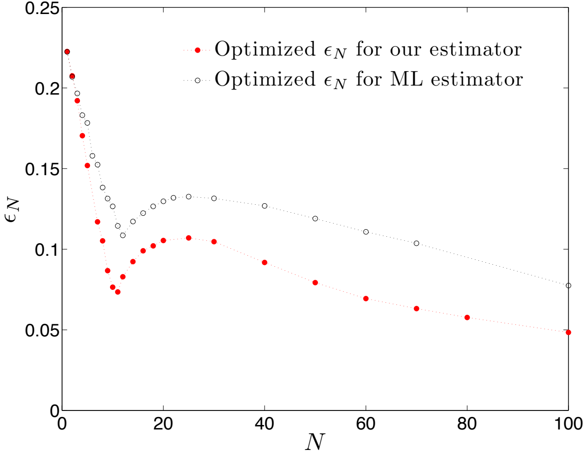
Qubit estimation: An act of desperation (2)

Extremal risks (= averaged squared Hilbert-Schmidt distances)



Qubit estimation: An act of desperation (3)

Optimal ϵ_N values



Quantum probabilities $\longrightarrow$ Quantum state (1)

For complete or over-complete data there is a unique statistical operator $\widehat{\rho}$ for each consistent ($=$ permissible) set $\widehat{p} = \{\widehat{p}_1, \widehat{p}_2, \ldots\}$ of estimated probabilities. As a rule, the conversion $\widehat{p} \to \widehat{\rho}$ is not a complicated procedure.

Example: $\widehat{\rho} = \sum_{k=1}^{4} \widehat{p}_k (6\Pi_k - 1)$ for qubit tomography with the tetrahedron measurement.

In the case of incomplete data, the estimated probabilities do not determine the state estimator uniquely. Additional conditions must be imposed if a unique $\widehat{\rho}$ is wanted.
One natural possibility is to invoke Jaynes's Principle:

Choose the $\widehat{\rho}$ with the largest entropy.

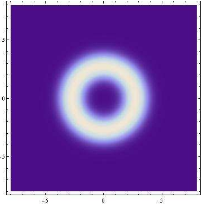
Quantum probabilities $\longrightarrow$ Quantum state (2)

In practice, however, another strategy is often applied: Truncate the Hilbert space, so that the data is (over-)complete in the restricted space and then find the unique $\widehat{\rho}$ there.
We disagree with this practice, because of its lack of justification and its ad-hoc nature, and much prefer the Jaynes-Principle procedure.
There could also be other acceptable procedures.

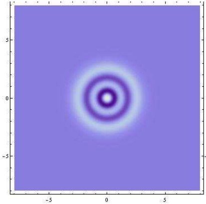
The implementation of the maximum-entropy method—in conjunction with MaxLik estimation, say—is possible with reasonable computational resources.

Quantum probabilities $\longrightarrow$ Quantum state (3)

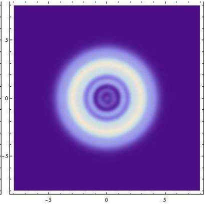
Truncated Hilbert space vs MaxLik-MaxEnt method



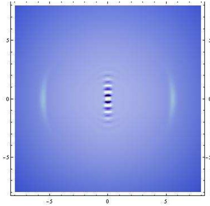
(a) True state



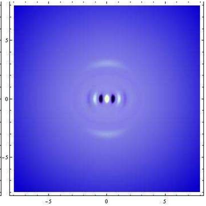
(b) 5-dimensional ML estimator



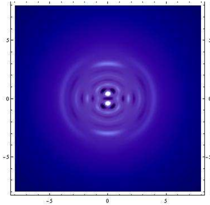
(c) 11-dimensional MLME estimator



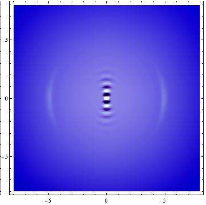
(a) True state



(b) 8-dimensional ML estimator

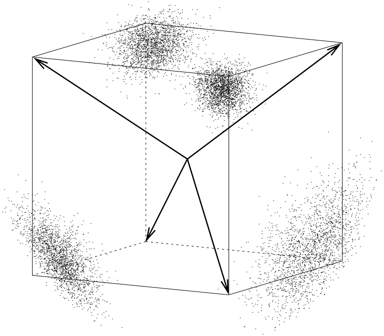


(c) 10-dimensional MLME estimator



(d) 15-dimensional MLME estimator

Outlook



From point estimators to region estimators: How?

References

arXiv:0906:3985, Phys. Rev. A **81** (2010) 052339

arXiv:0907.4258, Opt. Commun. **283** (2010) 724

arXiv:1008.1138, Phys. Rev. A **82** (2010) 042308

arXiv:1102.2662, Phys. Rev. Lett. **107** (2011) 020404

arXiv:1105.4561, Phys. Rev. A **84** (2011) 022327

arXiv:1110.1202, Phys. Rev. A **84** (2011) 062125

arXiv:1202.1713, Phys. Rev. A **85** (2012) 042317

arXiv:1202.5136, Int. J. Quant. Inf. **11** (2012) 1250038

arXiv:1207.0183
arXiv:1207.5386 two manuscripts submitted to NJP (special issue)

THANK YOU