ISCQI 2011: Jan 19 – 22

The Born Rule, the Frequency Operator and the Infinite Limit

R. Srikanth
Poornaprajna Institute, Bangalore- 560 080

With Prof. Dipankar Home, Bose Institute and
Prof. Somesh Kumar, IIT Kharagpur.

Outline of talk

(I) The quantum frequentist approach (Finkelstein 1963; Hartle 1968)

(II) Criticism of (I) by Farhi, Gutmann, Goldstone 1989; Gutmann 1995 – strong law version required.

(III) Criticism of (I) by Squires (1990): frequentist approach fails.

(IV) Criticism of (I, II) by Caves et al. (2004): frequentist approach fails.

(V) Modifying the frequentist approach to evade the above criticisms; but then resurrecting the criticism (IV) by modifying it.

(VI) Presenting a fundamental philosophical criticism of the quantum frequentist program.

A Postulate of Quantum Mechanics

The probability for obtaining outcome j in a measurement of observable $A = \sum_k \lambda_k \Pi_k$ is, by the **Born rule**,

$$\Pr(j) = \langle \Psi | \Pi_j | \Psi \rangle .$$

where Π_k is the projector onto the eigensubspace of A having eigenvalue λ_k.

Can one derive this from the remaining structure of quantum mechanics (QM)?

One possible route: Gleason's theorem (1958)

For dim ≥ 3, if probability of outcome $|j\rangle$ is non-contextual (independent of how the remaining part of the basis is completed), then

$$\mathrm{Prob}(j) = \mathrm{Tr}(\Pi_j \rho),$$

which is the Born rule applied to density operators.

Example: Given two bases:

$$A \equiv \{|0\rangle, |1\rangle, |2\rangle\} \text{ and } B \equiv \{|0\rangle, \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle)\}.$$

The prob($|0\rangle$) does not depend on whether A or B is used to complete the basis.

But then: we have merely traded one postulate (Born rule) for another (non-contextuality).

Non-contextuality from no-signaling?

One might ask: why then is Nature **non-contextual**?

Possible answer: Contextuality implies signaling (from previous example):

$$A \equiv \{|0\rangle, |1\rangle, |2\rangle\} \text{ and } B \equiv \{|0\rangle, \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle)\}.$$

Imagine $|0\rangle, |1\rangle, |2\rangle$ as spatially separate wavepackets, with B requiring a beam-splitter, whereas A not.

Therefore: No-signaling implies non-contextuality:

No-Signaling from WHNE?

Some might wish to ask: and why is Nature **no-signaling**?

One possible answer (RS, Physica Scripta 2010): “The World is Not Hard Enough” (WNHE)

WNHE: Hard problems should not be solvable efficiently, nor communication complexity trivialized (Van Dam 2005), using polynomial resources in the physical world.

Intuition here is that space is just another kind of information, and Nature is ultimately about computations performed on abstract bits of information located somewhere in Physical Reality (Consciousness ?!)

If signaling $\Rightarrow$ ~~WNHE~~, then assumption of WNHE would imply no-signaling.

Unfortunately, there exist *polynomial superluminal gates*, operations that lead to superluminal signaling but do not allow polynomial-time solution of NP-complete problems (RS Physica Scripta 2010).

In any case, the pattern is clear: trading one postulate for another does not help.

We arguably understand better only if fewer axioms are required to explain the theory or the new axioms are more intuitive (as pointed out by Pranaw yesterday!).

Another suggested route: the Frequency operator

The quantum analog of trying to **define** probability from properties of frequencies using the law of large numbers.

Bayesians claim that this doesn't work (classically) because the laws of large numbers are statements **within** probability.

The balance would have tipped in the frequentist's favor, if one could *derive* the Born rule from frequentist arguments.

The hope here is that: maybe inner product structure of QM gives additional leverage in the quantum as against classical case.

The Quantum Frequentist Program (QFP)

By studying infinitely many copies of a quantum system, **hopes** to eliminate references to probability from the postulates of quantum mechanics.

"In $|0\rangle$, the probability that a single X-measurement will yield $+1$ is $\frac{1}{2}$."

replaced by

"If the state is $|0\rangle^{\otimes\infty}$, $\frac{1}{2}$ of the X-measurements will yield $+1$."

Finite-copy frequency/average operator

(Finkelstein 1963; Hartle 1968).

$$
\begin{aligned}
F^N &= \sum_{n=0}^{N} \frac{n}{N} \Pi_j^N \\
&= \sum_{n=0}^{N} \frac{n}{N} \sum_{k_1,\cdots,k_N \in \{0,1\}} O_1^{(k_1)} \otimes \cdots \otimes O_N^{(k_N)} \delta(j, \sum_r \delta_{0,k_r}), \\
&= \frac{1}{N} \sum_{r=1}^{N} \sum_{k_1,\cdots,k_N \in \{0,1\}} O_1^{(k_1)} \otimes \cdots \otimes O_N^{(k_N)} \delta_{0,k_r}, \\
&= \frac{1}{N} \sum_{r=1}^{N} I^{\otimes r-1} \Pi_r I^{N-r}, \\
&= \frac{1}{N} \sum_{r=1}^{N} \Pi_r, \qquad (1)
\end{aligned}
$$

Consider the N-copy state $|\Psi_N\rangle \equiv |\psi\rangle^{\otimes N}$, where

$$|\psi\rangle \equiv c|0\rangle + s|1\rangle,$$

$c^2 + s^2 = 1$. It is straightforward to verify that

$$\begin{aligned} \|F^N|\Psi_N\rangle - c^2|\Psi_N\rangle\|^2 &= (c^4 - c^4/N + c^2/N) - 2c^4 + c^4 \\ &= \frac{c^2 s^2}{N} \end{aligned} \tag{2}$$

As $N \to \infty$, $|\Psi_N\rangle$ "nearly an eigenstate of F^N".

Axiom of Definite Outcomes (ADO)

ADO: If an observable O is measured on a system in an eigenstate $|\psi\rangle$ of O, i.e., $O|\psi\rangle = |\psi\rangle$, then the outcome is λ with certainty.

The interpretation apparently is that the Born rule follows from the Hilbert space structure since F^N finds Ψ_N to be an eigenstate with eigenvalue c^2, without invoking the Born rule!

Certain proponents of the Many-Worlds Interpretation of QM have tried to adapt QFP to that interpretation (Graham 1973; DeWitt 1973; Geroch 1984).

Squires' (1990) rejection of the QFP

Consider N copies of the state $|\Psi\rangle = c|0\rangle + s|1\rangle$, given by

$$|\Psi_N\rangle = |\Psi\rangle^{\otimes N}.$$

Assuming for simplicity that Nc^2 is an integer, $|\Psi^E\rangle$ is the symmetric superposition of the $K \equiv {}^N C_{Nc^2}$ states with precisely Nc^2 $|0\rangle$'s. Defining $\Pi^E \equiv |\Psi^E\rangle\langle\Psi^E|$, we find

$$\text{Prob(typical)} = ||\Pi^E|\Psi_N\rangle||^2 \equiv |\langle\Psi_N|\Psi^E\rangle|^2 = Kc^{2Nc^2}s^{2Ns^2} \equiv \zeta.$$

Applying Stirling's approximation one can show:

$$\lim_{N\to\infty} \zeta = 0.$$

Thus if the 'typical' eigenspace is orthogonal to $|\Psi_N\rangle$, then so is every other eigenspace of F^N (Squires 1990).

Farhi et al.'s (1989) criticism of F^∞

ADO applies only when the system is an exact eigenstate.

For exactness of eigenstate, we require infinite copies (in one shot, and not just convergence towards!) $\Rightarrow$ the strong law of large numbers.

The Finkelstein-Hartle result applies only to finite copies, governed by the weak law of large numbers.

Hence F^∞ must be defined using the strong law of large numbers.

Cantorian set theory

Orders of infinity (cardinality and ordinality).

A set that can be put in 1-to-1 relation with **N** has countably infinite elements (cardinality $\aleph_0$).

The set of reals in the interval $[0,1]$ is larger (provable via Cantor's diagonal argument) and has cardinality $2^{\aleph_0}$.

With cardinality of infinite sets, operations addition and subtraction do not change the cardinal type, but powering does.

A Hilbert space with a countable (resp. uncountable) orthonormal basis is called separable (resp. nonseparable).

Classical Intuition behind F^{∞}

Coleman and Lesniewski (1995) constructed a 'randomness operator' based on Kolmogorov-Chaitin complexity of a sequence of +1's and -1's. Their operator measures ? sequence of independent σ_x measurements yields random +1's and -1's. If the state is a product of σ_z eigenstates, they show the answer is a deterministic **yes** $\Rightarrow$ state an eigenstate.

Similarly, one can construct quantum operator that measures ? the outcome sequence of σ_x measurements has **any** given property.

Corresponding to the states, there exists classical probability measure on the generated $+1$'s and -1's. For many cases of interest, the classical probability that of set of sequences having the property, is 0 or 1 $\Rightarrow$ the state is an eigenstate of the operator with eigenvalue 0 or 1.

Construction of the infinite-copy frequency operator

Given a list of vectors $\{\Psi\} \equiv |\psi_1\rangle, |\psi_2\rangle, \cdots$, one defines their tensor product by

$$|\{\Psi\}\rangle \equiv \bigotimes_j |\psi_j\rangle.$$

Inner product:

$$\langle\{\phi\}|\{\psi\}\rangle = \Pi_{r=1}^{\infty}\langle\phi_r|\psi_r\rangle$$

Equivalence: $\{\phi\} \sim \{\psi\}$, if there exists $N \geq 1$ s.t

$$\Pi_{r=N}^{\infty}\langle\phi_r|\psi_r\rangle \geq 1-\epsilon.$$

i.e., equivalence if their tails are identical.

It can be shown that:
– the relation $\sim$ is an equivalence relation
– inequivalent vectors are orthogonal.

Component: subspace spanned by the infinite product vectors in a class.

Each component is separable: to see this, select sequence $\{\psi\} = |\psi_1\rangle, |\psi_2\rangle, \cdots$ from the eqivalence class that defines the component. Denote this component $\mathcal{H}^{\otimes\infty}_{\{\psi\}}$.

For each vector $|\psi_r\rangle$, choose orthonormal basis

$$|\psi_r, 0\rangle, \cdots, |\psi_r, D-1\rangle$$

where $|\psi_r, 0\rangle = |\psi_r\rangle$.

Define $\{i\} = i_1 i_2 \cdots$, where $i_k = 0, \cdots, D-1$, to be finite number of non-zero elements. These elements are countable. It can be shown that they span $\mathcal{H}^{\otimes\infty}_{\{\psi\}}$.

$\mathcal{H}^{\otimes\infty}$ is a nonseparable space (of dimension D^∞) formed by uncountably many tensor sums of *components*.

Infinite-copy frequency operaor

Assuming existence of limitng frequencies, we define frequency of a sequence $\{j\}$ to be:

$$f(\{j\}) = \lim_{N\to\infty} \frac{1}{N} \sum_{r=0}^{N} \left(\delta_{0 j_r} \right).$$

Following generalization of the strong law of large numbers (Feller 1971) required:

Let $X_n = \{0, 1\}$ be a sequence of RVs governed by probabilities $0 \leq q_r \leq 1$. Further, define the average probability:

$$f_{\{q\}} = \frac{1}{2} \left(\limsup_{N\to\infty} \frac{1}{N} \sum_{r=1}^{N} q_r + \liminf_{N\to\infty} \frac{1}{N} \sum_{r=1}^{N} q_r \right).$$

and define the frequency random variable by:

$$f^{\infty} = \frac{1}{2}\left(\limsup_{N\to\infty} \frac{1}{N}\sum_{r=1}^{N} \delta_{0X_r} + \liminf_{N\to\infty} \frac{1}{N}\sum_{r=1}^{N} \delta_{0X_r r}\right)$$

Then, by the Strong Law of Large Numbers (Feller 1965):

$$f_{\{q\}} = f^{\infty}$$

with probability $= 1$.

Consider component $\mathcal{H}_\Psi^{\otimes\infty}$ spanned by countable basis elements $|\Psi;\{i\}\rangle$ where sequences $\{i\}$ have only finite # non-zero entries. It is a subspace of non-separable $\mathcal{H}^{\otimes\infty}$.

Probability 'mass' associated with $|\Psi;\{i\}\rangle$ (to yield a sequence of outcomes) must come through a probability measure defined in $\mathcal{H}^{\otimes\infty}$.

Gutmann (1995): measure associated with all sequences beginning with $j_1,\cdots,j_N$ (meant to capture Born rule)

$$\begin{aligned}\nu_{|\Psi;\{i\}\rangle}(j_1,\cdots,j_N) &= \int d\nu_{|\Psi;\{i\}\rangle}(\{j'\})\Pi_{r=1}^N \delta_{j_r j'_r} \\ &= \Pi_{r=1}^N |\langle\Psi_r, i_r|B, j_r\rangle|^2.\end{aligned}$$

Within component, all sequences $\{i\}$ have same tails (consisting of 0's), product of terms $|\langle\psi_r|B, j_r\rangle|^2$.

A function of outcome sequences $\{j\}$ whose value is determined by the tail is independent of the first N values $j_1, \cdots, j_N$, for any N, is called a tail property (Feller 1971).

Therefore, when integrating a tail property over $d\nu_{|\psi;\{i\}\rangle}(\{j\})$, the measures for all $\{i\}$ will be the same.

Eg., a tail property is the average frequency:

$$\int d\nu_{|\psi;\{i\}\rangle}(\{j\})f(\{j\}) = f_{\{q\}},$$

where rhs is the average probability (for getting outcome 0) of a probability sequence $\{q_r\}$, where

$$q_r = \int d\nu_{|\psi;\{i\}\rangle}(\{j\})\delta_{0,r} = |\langle\psi_r, i_r|B, 0\rangle|^2 = q_{|\psi_r,i_r\rangle}(0).$$

As it is independent of $\{i\}$, we write:

$$f_{\{\psi\}} \equiv f_{\{q\}}.$$

(specifying only the base sequence of vectors).

Define projector onto frequency f:

$$||\Pi_f^\infty|\Psi;\{i\}\rangle||^2 \equiv \int d\nu_{|\Psi;\{i\}\rangle}(\{j\})\delta(f(\{j\}) - f).$$

The generalization of the strong law of large numbers quoted above (frequency of infinite measurments = avg. probability with prob = 1) reads as:

$$||\Pi_f^\infty|\Psi;\{i\}\rangle||^2 = \delta(f_{\{\psi\}} - f).$$

This determinism means the state in question is an eigenvector of the operator:

$$\Pi_f^\infty |\Psi; \{i\}\rangle = \delta(f_{\{\psi\}} - f)|\Psi; \{i\}\rangle. \tag{3}$$

Because this is based on a tail property, eigenvalue Eq. (3) is true for all vectors $|\Psi\rangle \in \mathcal{H}^\infty_{\{\psi\}}$.

By spectral decomposition theorem, we can define the infinite-copy frequency operator:

$$F^\infty|\Psi\rangle = f_{\{\psi\}}|\Psi\rangle.$$

In the present case, the component of interest has representative sequence $\{\psi\} = |\psi\rangle, |\psi\rangle, \cdots$ (infinite repetition state). So $q_r = |\langle\psi|B, 0\rangle|^2 = f_{\{\psi\}}$, the frequency associated with $\mathcal{H}^{\otimes\infty}_{\{\psi\}}$.

Caves et al. (2004) refutation of QFP even in its strengthened form

That there is a priori no special preference for the Born rule

$$q_r = |\langle\psi|B,0\rangle|^2 \Longrightarrow \mathcal{N}^{-1} g(|\langle\psi|B,0\rangle|),$$

with 'manually normalized' probabilities:

$$f_{\{\psi\}} = \frac{g(|\langle\psi|B,0\rangle|)}{\sum_j g(|\langle\psi|B,j\rangle|)}$$

Uncool, no doubt, but not disallowed by Nature !

$\Rightarrow$ no unique extension of the finite-copy frequency operator to the infinite-copy Hilbert space.

Weakening ADO

It turns out that the important criticism of Squires and Caves et al. can be evaded by suitably weakening ADO. The insight to do so comes from information theory.

Squires' expression for $|\langle\Psi_E|\Psi_N\rangle|^2$ is just the probability of obtaining a **typical** sequence from N independent tosses of a *classical* coin with probability distribution (c^2, s^2).

His result is an instance of *De Moivre* paradox: the number of heads tends to Nc^2, but the probability that it **exactly** Nc^2 is 0.

An easy way to understand this: the maximum of a binomial distribution tends to 0 as $N \to \infty$.

Classical information theory avoids this paradox by invoking ϵ-typicality instead of typicality.

For any $\delta, \epsilon > 0$, for N sufficiently large, each ϵ-typical sequence of symbols $x_1 x_2 x_3 \cdots x_n$ satisfies

$$2^{-n(H(X)+\epsilon)} \leq \text{Prob}(x_1 x_2 \cdots x_n) \leq 2^{-n(H(X)-\epsilon)}, \qquad (4)$$

and the total probability of all ϵ-typical sequences is greater than $1-\delta$. Here $H(X)$ is the entropy rate per symbol.

It can be understood also as follows: the distribution of the sample mean can be made as narrow as desired , but not 0 (unless strong law is invoked).

Analogously in the quantum case (Schumacher 1995): given n, ϵ, we define ϵ-typical subspace Λ as space spanned by vectors of $\rho^{\otimes n}$ with eigenvalue λ satisfying

$$2^{-n(S(\rho)-\epsilon)} \geq \lambda \geq 2^{-n(S(\rho)+\epsilon)}$$

Analogous to classical info, given $\delta, \epsilon > 0$ and n sufficiently large, the sum of ϵ-typical eigenvalues satisfies

$$\mathrm{Tr}(\rho^{\otimes n}\mathbf{E}) > 1 - \delta,$$

where $\mathbf{E}$ is projector to the typical subspace.

Weakening ADO.

Weak Axiom of Definite Outcomes (WADO): If an observable O is measured on a system that is ϵ-close (according to some criterion) to an eigenstate $|\psi\rangle$ of O, then the outcome is λ with probability greater than $1 - \delta(\epsilon)$.

One may argue that this doesn't uniquely "fix" quantum mechanics in that there is no unique way to weaken ADO.

However, it shows how to preserve "the spirit" of ADO while evading the objections of Squires and Caves et al.

With regard to Squires: The fact that $|\Psi_N\rangle$ is $\perp$ asymptotically to all eigenstates of F^N does not contradict the Weak QFP, because $|\Psi_N\rangle$ is nearly an eigenstate according to the ϵ-closeness criterion. we know that one weak criterion (e.g., the standard Schumacher one) that exists.

WADO implies that the Weak Law of Large Numbers suffices to define the frequency operator, and the Strong one is not appropriate to use. Thus the lack of uniqueness of extension of the finite-copy frequency operator to the infinite-copy Hilbert space is not a problem.

This does not mean that QFP can be salvaged, since a modified version of Caves et al.'s objection could be raised.

Lack of uniquess exists even in weak limit

Claim. Let $|\Psi_N^p\rangle \equiv |\psi\rangle^{\otimes N}$, where $|\psi\rangle \equiv a|0\rangle + b|1\rangle$ has been normalized according to the p-norm. Then there is a unique number, given by $q = |a|^p$, such that (Finkelstein-Hartle theorem)

$$\lim_{N\to\infty} ||F^N|\Psi_N^p\rangle - q|\Psi_N^p\rangle||^2 \equiv \lim_{N\to\infty} \Delta_N = 0. \qquad (5)$$

Proof.

$$\Delta_N^2 = (q - \langle\psi|P_{|0\rangle}|\psi\rangle)^2 + \frac{\langle\psi|P_{|0\rangle}|\psi\rangle(q - \langle\psi|P_{|0\rangle}|\psi\rangle)}{N}. \qquad (6)$$

With the choice of p-norm, we have $\langle\psi|P_{|0\rangle}|\psi\rangle = |a|^p$, and the theorem follows immediately. •

Thus the lack of uniqueness of the probability measure, pointed out by Caves et al. (2005), exists even without extending F^N to $\mathcal{H}^{\otimes\infty}$.

QFP demystified

Eq. 5 is simply a statement that variance of the binomial distribution (F_N represents N coin tosses and q the mean),

$$\lim_{N\to\infty} \Delta_N^2 = \frac{q(1-q)}{N} = 0. \tag{7}$$

Thus large-number or tail properties (which can be decribed without referring to the first n outcomes, for any finite n) are increasingly deterministic, and in the strong law limit, they are fully deterministic. The spirit of ADO is use this observation to define probabilities in terms of frequencies. But as we see, it does not fix q, which can be anything, incluing non-Gleasonian (contextual) probability laws.

Compatibility vs. implication

Thus, F-H theorem is a purely statistical result that relates single-copy probability with property of frequencies, and is devoid of any physical content regarding the form of the single-copy probability.

Thus the Born rule (2-norm) is **compatible**, but not a consequence of F-H theorem.

A Meta-theoretic objection to QFP

Any physical theory consists of a body of mathematical theory, and a meta-theoretic that supplies an interpretation.

The math of Quantum Theory could as well be applied to some other measure than probability, say 'brightness' !

Viewing FHP in another light: the 'bosonic binomial state'–

$$|\Phi_N\rangle \equiv \left(\sqrt{p}|0\rangle + \sqrt{1-p}|1\rangle\right)^{\otimes N} = \sum_{j=1}^{N} \sqrt{\left(\begin{array}{c} N \\ j \end{array}\right)} |(N,j)\rangle,$$

where $|(N,j)\rangle$ is the symmetrized state over all "typical sequences" of Np 0's and $N(1-p)$ 1's and p is assumed to be real.

As $N \to \infty$, most untypical kets will "fade away". But this does not mean, typical sequences are more **probable**, but instead just **brighter** !!

Thus Quantum Frequentist Program is as bankrupt philosophically, as mathematically!

We may yet find a deeper explanation for the Born rule, but not here.

Thanks to Himanshu Sharma and Ebad Kamil for discssions!

Thank You for your kind attention!