# Quantum Probability: Conditioning and Inference

**R. Scozzafava**
Università La Sapienza, Roma
Dipartimento di Metodi e Modelli Matematici
Via Scarpa 16, 00161 Roma, Italy.
e-mail: *romscozz@dmmm.uniroma1.it*

## Abstract

Looking at *conditional* probability from the right perspective allows one to avoid classic paradoxical aspects of quantum mechanics. We discuss, as an elementary analogue of quantum states, a classical example (an urn of unknown composition) with similar (putative) paradoxical situations, despite the fact that there is no quantum effect to which they could be ascribed. A crucial feature is that we look at them from an epistemic point of view and not from the viewpoint of a (seemingly) ontological reality.

**Keywords:** Conditional probability, Coherence, Quantum Mechanics

## 1 Introduction

Classical mechanics presupposes that a physical system can in principle be described by referring to the notion of mass point, or particle. Until the advent of quantum physics, looking at a physical system as a set of particles has never been questioned. Accordingly, to correlate instantaneous positions $X$ of a given particle with instants of times, a description of the form $X = f(t)$ appears as natural, and should have meaning even when the functional relation cannot be established, due to practical and accidental difficulties in the procedure of measuring $X$ as a function of $t$.

The following simple example shows that classical description must be abandoned not because of experimental difficulties, but because its use contradicts known laws of science. Consider the oscillating mass point constituted by the bob of a pendulum, and assume that the bob oscillates $10^{10}$ times per second. Since it is a fact of atomic physics that visible light (used as medium of report) requires about $10^{-8}$ seconds to be emitted or reflected, then the light–emitting mass would have to remain in a given position for approximately that length of time. In the present instance, however, the bob executes 100 vibrations within this period. So the classical correspondence has to be ultimately abandoned because its use would contradict the laws of optics. The same example suggests the way–out: notice that even a snapshot taken by a camera is not able to show the correspondence between $X$ and $t$, nevertheless it would give essentially a correlation between the time the bob spends within a given interval $dx$ and the location of that interval. In other words, we get a relationship between $x$ and the probability $p(x)\,dx$ of encountering the bob in $dx$.

This kind of description is characteristic of quantum mechanics, which aims at deducing such probability relations in a logically consistent fashion. As a consequence, this approach has left the concept of particle very ill defined and loosing its physical significance. So from this situation there has arisen a claim that quantum mechanics leads to a dualism, to the strange conception that fundamental entities of physics like electrons are both particles and waves, with ensuing paradoxical conclu-

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 923–929
Torremolinos (Málaga), June 22–27, 2008

sion and faulty agreement between theory and empirical facts.

In this paper we emphasize the role of *conditional* probability, and we show that looking at it from the right perspective allows one to avoid classic paradoxical aspects of quantum mechanics. Moreover, we discuss, as an elementary analogue of quantum states, a classical example (an urn of unknown composition) with similar (putative) paradoxical situations, despite the fact that there is no quantum effect to which they could be ascribed. In fact a great number of phenomena (even if not typical of the way of thinking required by quantum theory) that seem mysterious from the viewpoint of a (seemingly) ontological reality, look instead as natural from an epistemic point of view.

## 2 Probability and frequency

Probability is a general tool guiding induction and *must not be restricted only to the empirical interpretation* coming from its *evaluation* by an observed frequency. It is true that in many cases the *value* of a probability can be sensibly expressed by a suitable frequency, but this does not entail that the latter can be taken as *definition* of probability. A frequency observed in repeated trials is a fact, not a probability.

Our discussion has a close connection with a famous statement – *"Probability does not exist"* – put at the very beginning of his book[3] by B. de Finetti (1974, Preface): its provocative and striking flavor is a criticism of any attempt to represent (or interpret) probabilities as physically real things, similar to quantities like mass or length, existing in the world and independent of us. In fact, pursuing this myth of objectivity leads to "define" probability as "limit" of the frequency when the number $n$ of trials (performed, as it is said, "under the same conditions") increases (by the way, if the trials were really performed under **exactly** the same conditions, they would always produce the same result!).

But notice that we have the following alternatives:

($\alpha$) Either the "limit" frequency does not exist,

or it exists (and so, beyond a suitable $n_o$ – which is, by the way, unknown! – the observed frequency coincides with the "limit" frequency within an arbitrarily preset level $\epsilon > 0$ of approximation) and we can then distinguish two situations:

($\beta_1$) either the number of observed trials is not sufficient (smaller than $n_o$),

($\beta_2$) or these trials are enough (at least $n_o$).

Clearly, in the first two cases ("limit" frequency nonexistent or number of trials less than $n_o$) we cannot conclude anything. To make significant the conclusions that could be drawn in the case ($\beta_2$), we need assuming that this case is "much more probable" than the other two (and the only meaning of "probable" can be the subjective one as degree of belief). But the latter is just the meaning that one would – by resorting to the frequency evaluation – aim at avoiding ...!

Therefore a strict frequentist view, as that often considered in statistical physics and in quantum mechanics, cannot avoid a subjective framework; moreover, it ends up by transferring the reasoning from practically verifiable events (such as a frequency on a **finite number** of trials) into fictitious entities, practically out of control and absolutely indeterminate. In other words, it needs resorting to two unjustifiable (and inconclusive) arguments: firstly from a finite framework to an infinite one (the putative existence of the "limit" frequency); and then from an infinite framework to a finite one (as an approximation for "large" $n$).

## 3 "Small" probabilities

Even if it is true that in many cases the value of a probability is "very near" to a suitable frequency, in every situation in which something "very probable" is looked on as "practically certain", there are "small" probabilities that are actually ignored, so making illegitimate also any probabilistic interpretation of physical laws. For example, a probabilistic explanation of the diffusion of heat *must take into account the fact that the heat could accidentally move from a cold body to a warmer one*, making the former even colder and the latter even warmer. This fact is very improbable only because the "unordered" configurations (*i.e.*, heat equally diffused) are far more numerous than the "ordered" ones (*i.e.*, all the heat in one direction), and not because unordered configurations enjoy some special status.

As a second example, consider the following (very simple) one: when we press "at random" 18 keys on a keyboard, we would like to explain why the statement *"to be or not to be"* has not been written; on the other hand, while forecasting the occurrence of any one of all other sequences, we cannot consider it **impossible** that *"to be or not to be"* could come out. In fact, if we were arguing in this way, it would mean also denying the possibility of explaining why we got just that sequence which we actually got – since *it had the same probability* of being typed as that piece of "Hamlet". So, why it is so difficult to see that piece by Shakespeare coming out – or else: to see water freezing on the fire – even in a long series of repetitions of the relevant procedure? It is just because the relevant "waiting times" (inversely proportional to the corresponding probabilities) are extremely large.

Notice that the difference between an *impossible* fact and a *possible* one – also with a very small probability, or even zero (it is well-known that we may have "many" *possible* events with zero probability) – is really enormous, since it is not a matter of a *numerical* difference, but of a *qualitative* (*i.e.*, logical) one. In conclusion: a probabilistic law cannot be **falsified** (in the sense of Popper); it may *possibly* be modified, if one thinks that the new one is a better *model* (a sort of "updating", in the sense of Bayesian inference).

We could say that a "law" which is based on a probabilistic interpretation **does not** play the role of forecasting that **a fact will occur**, but rather that of explaining **why we forecast that this fact will (with high probability) occur**.

## 4 Conditional probability and coherence

Before proceeding further in the discussion, it is essential to make a digression on probability theory through the concept of **coherence**[3].

Coherence allows you to assess your probability for as many or as few events as you feel able and interested, possibly then going on by extending it to further events. This has many important theoretical and applied consequences: for example, the axiomatic counterpart of de Finetti's theory is weaker than the traditional Kolmogorov's approach and makes simpler and more effective the "operational" aspects.

An *event* can be singled-out by a (nonambiguous) *proposition* $E$, that is a statement that can be either *true* or *false* (corresponding to the two "values" 1 or 0). Since in general it is not known whether $E$ is true or not, we are *uncertain* on $E$. Classical examples of events are: *(i)* any proposition that describes the so-called "favorable" cases to a possible outcome $E$, a situation which is typical in the well known classical (or combinatorial) approach to probability; *(ii)* given a (finite) sequence of trials performed "under similar conditions", any proposition describing a possible result occurring in each trial (frequentist approach); but notice that an *event* is also **anything else that can be expressed by a sensible proposition**.

Probability is looked upon as an "ersatz" for the lack of information on the actual "value" (true or false) of the event $E$, and it can be regarded as a measure of the *degree of belief* in $E$ held by the *subject* that is making the assessment. In particular, the two most popular approaches to probability may be taken just as useful *methods of evaluation* (when we judge that a suitable "symmetry" exists allowing an assessment based on combinatorial considerations, or that the different trials needed for a frequentist assessment are performed under "similar" conditions). Notice that an uncertain event $E$ may become true (e.g., statistical data), so reducing to the *certain event* $\Omega$ , or become false (when its *contrary* $E^c$ is true), so reducing to the *impossible event* $\emptyset$.

In general, it is not enough directing attention just toward an event $E$ in order to assess "convincingly" its probability, but it is also essential taking into account *other* events which may possibly contribute in determining the "information" on $E$. Then the fundamental tool is *conditional probability*, since the true problem is not that of assessing $P(E)$, but rather that of assessing $P(E|H)$, taking into account all the relevant information carried by some other event $H$ (possibly corresponding to statistical data, necessarily regarded as "assumed" and not as "acquired": for a clarification of these two terms, see Sect.5).

**Definition** – Let $\mathcal{E}$ and $\mathcal{H}$ be two families of events, with $\emptyset \notin \mathcal{H}$, where $\mathcal{E}$ is an algebra and $\mathcal{H} \cup \{\emptyset\}$ is a subalgebra of $\mathcal{E}$. A nonnegative function $P(\cdot|\cdot)$ is said a (finitely additive) *conditional probability* on $\mathcal{E} \times \mathcal{H}$ if

*(a)* for any given event $H \in \mathcal{H}$ and $n$ mutually exclusive events $A_1, ..., A_n \in \mathcal{E}$, the function $P(\cdot|H)$, defined on $\mathcal{E}$, satisfies

$$P((\bigcup_{k=1}^{n} A_k)|H) = \sum_{k=1}^{n} P(A_k|H)\ ,\ P(\Omega|H) = 1\,;$$

*(b)* $P(H|H) = 1$ for any $H \in \mathcal{H}$ ;
*(c)* given $E, H, A$ such that $E \in \mathcal{E}$, $A \in \mathcal{E}$, $H \in \mathcal{H}$, then

$$P(E \cap A|H) = P(E|H)P(A|E \cap H)\,.$$

In particular, choosing $\mathcal{H} = \{\Omega\}$ and putting $P(E|\Omega) = P(E)$ for any $E \in \mathcal{E}$, the function $P$ is said a *probability* if condition *(a)* holds. Notice also that *(c)* reduces, when $H = \Omega$, to the classic product rule for probability.

From *(a)* and *(c)* it follows easily, for any two events $E$ and $H$:

$$P(E) = P(H)P(E|H) + P(H^c)P(E|H^c)\,,$$

or, more generally, given any partition $\{H_r\,,\ r = 1, 2, ..., n\}$ of the certain event

$$(*) \qquad P(E) = \sum_{r=1}^{n} P(H_r)P(E|H_r)\,.$$

We consider now an *arbitrary* set $\mathcal{C}$ of conditional events, with no underlying structure. A function $P$ on $\mathcal{C}$ is called **coherent** *if and only if* $P$ is the restriction of a conditional probability on $\mathcal{E} \times \mathcal{H} \supseteq \mathcal{C}$. Then $P$ may be called a *coherent conditional probability* (a weaker notion than that of conditional probability).

**Extension Theorem** (de Finetti[2]) – If $\mathcal{C}$ is an arbitrary family of conditional events and $P$ a corresponding assessment on $\mathcal{C}$, then there exists a (possibly not unique) coherent extension of $P$ to *any* arbitrary family $\mathcal{G}$ of conditional events, with $\mathcal{G} \supseteq \mathcal{C}$, if and only if $P$ is coherent on $\mathcal{C}$.

Notice that, since $P$ can be *directly* introduced as a function whose domain is an arbitrary set of conditional events, bounded to *satisfy only the requirement of coherence*, $P(E|H)$ can be assessed and makes sense for any pair of events $E$, $H$, with $H \neq \emptyset$, and, moreover, the knowledge (or the assessment) of the "joint" and "marginals" probabilities $P(E \cap H)$ and $P(H)$ is not required. In particular, there is no need, as in the usual approach – where the conditional probability $P(E|H)$ is introduced by *definition* as the ratio between the probabilities $P(E \cap H)$ and $P(H)$ – of assuming positive probability for the given conditioning event.

If we refer just to a *single* event, its probability *can* be assessed by an observed frequency in the past (since a frequency is a number between 0 and 1, and this is a *necessary and sufficient* condition for coherence when *only a single event* is considered). But things are not so easy when more than one event (conditional or not) is involved, since consistency problems must then be taken into account.

For a thorough treatment of coherent probabilities, see the book[1].

## 5 Conditioning and quantum preparation

Property (c) of conditional probability is crucial in order to fully exploit its inferential meaning. In fact, what is usually (in the relevant literature) emphasized is only the role played by property *(a)* – *i.e.*, a conditional probability $P(\cdot|H)$ **is a probability for any**

**given** $H \in \mathcal{H}$.

For example, in the jargon of quantum probability, the set $\mathcal{H}$ is called the family of **preparations** and it is the physical counterpart of the notion of "conditioning": a probability $P(E|H)$ is interpreted as the (approximate) relative frequency of the event $E$ in an ensemble of systems prepared in such a way that the event $H$ is **certainly verified** for each of them.

This is a very restrictive (and misleading) view of conditional probability, *corresponding trivially to just a modification of the so-called "sample space"* $\Omega$. It is instead essential – for a correct handling of the subtle and delicate problems concerning the use of conditional probability – to regard as a "variable" also the *conditioning* event $H$. In other words, the "status" of $H$ in $E|H$ is not just that of something representing a given **fact**, but that of an (uncertain) **event** (like $E$) for which *the knowledge of its truth value is not necessarily required*.

In order to see the problem from a different perspective (and also in a context which has nothing to do with statistical physics or quantum mechanics), we discuss some fundamental remarks concerning, for a given conditional event $E|H$, the "information" represented by $H$: the main point is a distinction between "assumed" and "acquired" $H$.

For example, in Bayesian inferential statistics, given any event $E$, with *prior* probability $P(E)$, and a set of events $E_1, ..., E_n$ representing the possible statistical observations, with likelihoods $P(E_1|E), ..., P(E_n|E)$, all *posterior* probabilities $P(E|E_1), ..., P(E|E_n)$ are usually pre-assessed through Bayes' theorem (which, by the way, is a trivial consequence of conditional probability rules). In doing so, each $E_k$ $(k = 1, ..., n)$ is clearly regarded as "assumed". If an $E_k$ occurs, $P(E|E_k)$ is chosen – among the prearranged posteriors – as the updated probability of $E$: *this is the only role played by the "acquired" information* $E_k$ (the sample space is not changed!).

In other words, the above procedure corresponds, putting $P(E|E_k) = p$, to regard a conditional event $E|H$ **as a whole** and interpret $p$ as "the probability of ($E$ given $H$)" and not as "(the probability of $E$), given $H$". In fact, the latter interpretation is unsustainable, since it would literally mean "if $H$ occurs, then the probability of $E$ is $p$", which is actually a form of a logical deduction leading to absurd conclusions, as in the following simple example.

Consider a set of five balls $\{1, 2, 3, 4, 5\}$ and the probability of the event $E$ that a number drawn from this set *at random* is even (which is obviously 2/5): this probability could instead be erroneously assessed (for instance) equal to 1/3, since this is the value of the probability of $E$ conditionally on the *occurrence* of each one of the events $E_1 = \{1, 2, 3\}$ or $E_2 = \{3, 4, 5\}$, and one (possibly both) of them *will certainly occur*.

## 6 Urn of unknown composition

Take a box with a given number $N$ of balls: each one is either white or black, but *the actual composition of the box* (*i.e.*, the number $r$ of white balls and hence that $N - r$ of black ones) *is not known*. Consider the random experiment consisting in drawing one ball from it: how to assign a probability to the event (denoted by $E$) *"the ball drawn from the box is white"*? A strict supporter of probability measured **only** by means of an observed frequency, could make the choice of simply ignoring the fact that the composition of the box is unknown: given a "sufficiently large" natural number $n_o$ , he could perform the experiment consisting of $n_o$ drawings with replacement from the box and evaluate (*a posteriori*) the required probability by the ratio between the observed number $X$ of white balls and the number $n_o$ of drawings. But he is not willing to assign any *a priori* probability to $E$: for him this probability **does not exist**.

However, since probabilities exist inasmuch as one creates them as useful substitutes for *a lack of information about something*, no "physical" or combinatorial interpretation is needed. Given any set of events whatsoever, coherence essentially imposes on the probabilities that may be assigned to them the only restriction that they must not be in contradiction amongst themselves, according to the aforementioned syntactic rules.

So in the example under consideration we may assign a probability distribution to the possible compositions of the box, *i.e.* to the $N+1$ events *"there are r white balls in the box"*, with $r = 0, 1, 2, ..., N$, and so, denoting by $H_r$ these events and introducing suitable conditional probabilities $P(E|H_r)$, the probability of $E$ can be represented by

$$(1) \qquad P(E) = \sum_{r=0}^{N} P(H_r)P(E|H_r) \,.$$

In particular, we may think (for example) as if the given box were chosen at random among $N+1$ boxes corresponding to all possible compositions, so that the probabilities $P(H_r)$ (summing up to 1) are equal: then formula (1) takes the simpler form

$$P(E) = \sum_{r=0}^{N} \frac{1}{N+1} P(E|H_r) \,.$$

On the other hand, evaluating, for each given $r$,

$$(2) \qquad P(E|H_r) = \frac{r}{N} \,,$$

we may conclude, by straightforward computations, that

$$(3) \qquad P(E) = \frac{1}{2} \,.$$

Notice that the previous frequentistic evaluation of the probability of the event $E$ by means of experiments producing the ratio $X/n_o$ is in fact an evaluation of the *conditional* probability $P(E|H_r)$ corresponding to the *given but unknown* $r$: so "in the long run" ($n_o$ "sufficiently large") the *observed* frequency takes (approximately) the value (2). Instead the value (3), which is a perfectly legitimate and sensible evaluation of the probability of $E$, should be regarded as unacceptable from the point of view of a *strict* frequency interpretation of probability: in fact the drawings (with replacement) from the given box will almost certainly give a frequency near to the value (2), and so different from 1/2 (except in the particular case in which the number $r$ of white balls was about half of the total number $N$ of balls).

The conclusion is that a frequentist evaluation of $P(E)$ leads to a violation of the (syntactic) rule (1). Even if there is no quantum effect to which the lack of validity of formula (1) could be ascribed, we can picture the situation using the jargon of quantum physics: the ball we draw from the urn is a *system*, the possible compositions of the urn (including the relevant probabilities) correspond to its *state vector*, and its colour is an *observable*.

Notice that the observation of $E$ (white ball) entails (by Bayes' theorem) an updating of the probabilities $P(H_r)$, and so also this inferential aspect must be taken into account for the subsequent observations. In fact this information changes the state of the system.

## 7 Are observed frequencies coherent extensions?

The lesson we have learned (from the discussion of the previous section) is – again – that probabilities cannot be identified with observed frequencies. Moreover, these frequencies correspond (depending on the – unknown – value of $r$) to different and incompatible experiments, each one referring to a different (ideal) box. Nevertheless these experiments are not (so to say) "mentally" incompatible if we argue in terms of the general (epistemic) interpretation of probability: then, for a coherent evaluation of $P(E)$ *we must* necessarily rely *only* on the above value obtained by resorting to the second member of (1), even if such probability does not express any sort of "physical property" of the given box. In other words, the previous discussion can be seen as an instance of the problem of *finding a coherent extension of some beforehand given probabilities* (see the Extension Theorem of Sect.4). Interpreting $E$ as $E|\Omega$ and $H_r$ as $H_r|\Omega$, it is easily seen that the value $P(E)$ given by (1) **is** a coherent extension of the conditional probabilities $P(E|H_r)$ and $P(H_r|\Omega)$, while in general a value of $P(E)$ obtained by measuring a relevant frequency **may not**. In other words: while a convex combination of conditional *probabilities* can be a *probability*, a convex combination of conditional *frequencies* is not necessarily apt to evaluate a particular probability ...

Therefore, if in some applications (quantum mechanics, drawings balls from boxes, or anything else) one introduces a "probabilistic" model whose corresponding function $P$ fails to satisfy the relevant syntactic rules, then necessarily $P$ is not coherent, and so it should be clear that any attempt at accepting $P$ as a sort of "new kind of probability" necessarily amounts to deny and getting rid of the **very concept** of conditional probability.

## 8 Conclusions

In quantum mechanical experiments, the identification of (conditional) probabilities with some statistical data (essentially, observed frequencies) may lead to results which are in contradiction with other experiments (still involving observed or expected frequencies).

The classical two-slit experiment, discussed from a probabilistic point of view by Feynman[4], is an interesting illustration of the quantum mechanical way of computing the relevant probabilities: our interpretation in term of coherent probability has been already discussed in [5, 6]. The situation looks very similar to what happens with Bell's inequality. Without going into details, we just recall that Bell calculates the expectation values for certain physical quantities, and then shows that these quantum mechanical expectations violate Bell's inequality. As Stapp[8] puts it (even if not in a completely clear way), quantum states "are to be interpreted as symbolic devices that scientist uses to make predictions about what they will observe under specific conditions", and else "Bell's theorem proves that the ordinary concept of reality is incompatible with the statistical predictions of quantum theory".

In conclusion, referring to repeated experiments, much care is needed when dealing with frequencies and probabilities: making computations of a conditional probability by the frequency relevant to a given experiment involves (so to say) a choice, in the sense that it is no more allowed to consider – in the same framework – **other** experiments, otherwise the rules of conditional probability (in particular, those referring to a "variable" **conditioning** event) could be violated.

## References


[1] G. Coletti and R. Scozzafava (2002), *Probabilistic Logic in a Coherent Setting.* Kluwer Academic Publishers, Trends in Logic, n.15, Dordrecht - Boston - London.

[2] B. de Finetti (1949), "Sull'impostazione assiomatica del calcolo delle probabilità", *Annali Univ. Trieste*, 19, 3–55 – Engl.transl.: Ch.5 in *Probability, Induction, Statistics*, Wiley, London, 1972.

[3] B. de Finetti (1974), *Teoria della probabilità*, Einaudi, Torino (1970) – Engl. transl.: *Theory of Probability*, Voll.1 and 2. Wiley, Chichester.

[4] R. Feynman (1951), "The concept of probability in quantum mechanics", *Proc. 2nd Berkeley Symp. on Mathematical Statistics and Probability*, University of California Press, Berkeley, 533–541 .

[5] R. Scozzafava (1991), "A classical analogue of the two-slit model of quantum probability", *Pure Mathematics and Applications, Series C*, 2, 223–235.

[6] R. Scozzafava (2000), "The role of probability in statistical physics", *Transport Theory and Statistical Physics*, 29, 107–123.

[7] R.W. Spekkens (2007), "Evidence for the epistemic view of quantum states: a toy theory", *Physical Review* A 75, 032110.

[8] H.P. Stapp (1972), "The Copenhagen interpretation", *America Journal of Physics*, 40, 1098–1116.