THE ROLE OF PROBABILITY IN STATISTICAL PHYSICS

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

In this expository paper we discuss the role of probability in statistical physics: it should be that of 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. Moreover, in each situation in which one thinks that probabilities are physically real things and believes that something which is only "very probable" can be regarded as being "practically certain", some "small" probabilities are actually ignored: this attitude of mind may render illegitimate also the probabilistic interpretation of the relevant physical laws. A "law" which is based on a probabilistic interpretation cannot aim at explaining why a fact that we forecast will occur, but rather at explaining why we forecast that this fact will occur. Finally, we emphasize the role of *conditional* probability, and we show that looking at it from the right perspective also allows one to avoid classic paradoxical aspects of quantum mechanics, without the need to resort to the interpretation based on the dualism "particle – wave".

1 From Boltzmann to de Finetti

The subject dealt with in this expository paper refers to a field in which many difficult and subtle issues intertwine each other: so an effort to succeed in pointing out and isolating what is the real core of probabilistic arguments in physics is necessary. We will give up any discussion of specific *intricate* questions, in favour of a deepening of the *intriguing* side of the problem.

Even if physicists became to think to nondeterminism around 1925-27 ("modern" quantum theory), statistical physics made its first steps in the 1850's, with the formulation of the theory of heat.

Maxwell's name is associated to the "ensemble" view of probability in statistical physics, but this approach was actually invented by Boltzmann (though he is not always credited for it) and further developed by Gibbs. Boltzmann advocated a **time-average** interpretation of probability, introducing the concept of **physical state**, as given by a probability distribution.

This time-average notion of probability is closely related to a "statistical" or "frequentist" probability, and Boltzmann claims that "the second law can never be proved mathematically from the equation of dynamics alone" and that it is a theorem of probability: so the observed regularity in the behavior of gases is a statistical effect from a large number of molecules – sometimes dubbed as "statistical determinism" (but – as we shall deepen in the sequel – the role of probability is not to establish which previsions are "right", but which previsions are "sensible" according to our present knowledge).

Even the terminology may reflect the attempt to escape an actual dealing with "uncertainty", by an attitude toward a (putatively ...) "objective" view of Science.

So, 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!)

Let us just mention (we will not dwell on these aspects) that it is possible – through the concept of **exchangeability** – to **evaluate** a probability taking into account suitable observed frequencies (relative to just a finite number of trials) also in the framework of "coherence" – the leading concept of the so-called "de Finetti's subjective approach"¹, where the *meaning* of the probability of an event is that of "degree of belief" in the verifiability of the event (independently of its *method of evaluation*).

2 A general view of probability through coherence

An important issue of the previous discussion concerns the careful distinction that must be made between the *meaning* of probability and its *methods of evaluation*: ignoring this distinction would be analogous to identifying the concept of temperature with the number shown by a thermometer, so being not entitled to speak of temperature in a room without a thermometer. This confusion often appears not only in statistical physics: for example, we will show that (putative) paradoxes similar to the classic ones arising in the statistical description of quantum phenomena are not typical of the way of thinking required by quantum theory.

Consider also the following example: we can test a given object to find out at what temperature it melts or under what pressure it splits, but we clearly cannot carry out both these tests. This may be misinterpreted as a paradoxical situation only if we pretend to express it by saying loosely that "it is meaningful only to ask whether the given object is either refractory or robust, but not to ask these questions together". In other words, though the critical cases of quantum theory have nothing essentially different from the latter example, they look like paradoxical only when one refers to concepts and situations for which it seems "more natural" to expect an answer concerning some "composite" question (as it was for position and momentum before the advent of Heisenberg's uncertainty relation).

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

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 Emay become true (e.g., statistical data), so reducing to the *certain event* Ω , 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.6).

Dealing with conditional probability requires the introduction of *condi*tional events E|H, with $H \neq \emptyset$. The usual interpretation in terms of a betting scheme is the following: if an amount p – taken as the value of P(E|H) – is paid to bet on E|H, we get, when H is true, an amount 1 if also E is true (the bet is won) and an amount 0 if E is false (the bet is lost), and we get back the amount p if H turns out to be false (the bet is called off). In short, the (random) value taken by E|H is just the amount got back when one bets on it by paying an amount p. In particular, an (unconditional) event E can take only two values (1 or 0) and can be looked on as $E|\Omega$, where Ω denotes the *certain* event. (It is also possible to consider a "scale factor" λ , *i.e.* to pay $p\lambda$ and receive – when the bet is won – an amount λ : we were referring above to the case $\lambda = 1$). For a deeper discussion of these foundational aspects, see the expository paper².

The point of defining probability (conditional or not) in terms of bets is to give it an *unmistakable, concrete and operational meaning, valid for any kind of event*, not only those corresponding to the classic combinatorial or frequentist evaluations. You may assess its probability **however you like**: obviously, you are not allowed to violate the relevant *syntactic* rules! These come out from the following framework: let C be an arbitrary family of conditional events and P a real function defined on C. Given any finite subfamily

$$\mathcal{F} = \{E_1 | H_1, \dots, E_n | H_n\} \subseteq \mathcal{C},\$$

we put $P(E_i|H_i) = p_i$ for i = 1, ..., n. Then, denoting by b the *indicator* function of an event B, we consider the random quantity

$$G = \sum_{i=1}^{n} \lambda_i h_i (e_i - p_i) \, ,$$

which can be interpreted as the gain corresponding to a combination of n bets of amounts $p_1\lambda_1, \ldots, p_n\lambda_n$ on $E_1|H_1, \ldots, E_n|H_n$, with arbitrary real stakes $\lambda_1, \ldots, \lambda_n$. Denoting by H_0 the union $H_1 \cup \ldots \cup H_n$ and by $G_{|H_0}$ the restriction of G to H_0 , we have the following

Definition – The real function P on C is *coherent* if, given the assessment

 $\mathcal{P} = (p_1, \ldots, p_n)$ on each finite family $\mathcal{F} \subseteq \mathcal{C}$, with $p_i = P(E_i|H_i)$, for every choice of the real numbers $\lambda_1, \ldots, \lambda_n$ the possible values of the corresponding gain $G_{|H_0}$ are neither all positive nor all negative.

Notice that this view is based on *hypothetical* bets: the force of the argument does not depend on whether or not one actually has the possibility or intends to bet. In fact a method of assessing probabilities making one a sure loser or winner *if he had to gamble* (whether or not he really will act so) would be suspicious and unreliable *for any purposes whatsoever*.

Remark – As we will discuss in Sect.5, coherence of P implies that it satisfies the axioms of a conditional probability: but the introduction of these syntactic rules through coherence allows a "wider" semantic interpretation of P as a (subjective) "degree of belief" (possibly, but not necessarily, evaluated by a suitable frequency).

3 Probability in Statistical Physics

It is important to underline that a strict frequentist approach (in the attempt to "define" – and so to *identify* – probability with the "limit" of a frequency) *does not avoid the subjective aspects, but just* … *hides them*. In fact the actual frequency observations can refer only to a **finite number** of trials, and so the introduction of the frequency as a "limit" is made under the assumption that this limit should be practically equal to the observed frequency: yet this circumstance cannot be considered "certain", but only "a highly probable" event, where the meaning of "probable" is necessarily nothing else than that based on a *degree of belief*, as in the subjective approach. But the latter is just the meaning that one would – by resorting to the frequency evaluation – aim at avoiding …!

Moreover, all the properties (and the relevant formal calculations) concerning the "limit" frequency (identified with the probability) can be interpreted in real situations only by carrying out the reverse procedure, *i.e.* through the approximation of the "limit" frequency by the frequency relative to a "very large number" of trials (actually assuming as "highly probable" – again – the validity of this approximation).

In other words, we have the following alternatives:

(α) 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:

 (β_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 (β_2), we need assuming that this case is "much more probable" than the other two (and the only meaning of "probable" can be – again – the most general one as degree of belief).

Therefore a strict frequentist view, as that often accepted in Statistical Physics, 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).

A possible objection to the above argument concerning the interpretation of a strict frequentist view of probability, could be the following.

From an empirical point of view, every highly probable event should be looked on as being practically certain, and probability theory should constitute an ideal scheme, where the agreement with empirical observations may be found only in an approximate way.

In other words, in every experimental science there is a theory that says how facts should occur, and an empirical postulate which states that they actually happened approximately that way.

But this seeming analogy is devoided of any value, since there would be analogy only if probability theory had some theorem stating, for example, that – the two outcomes being equiprobable – 500 out of 1000 tosses of a coin will show "head" and 500 "tail", while in practical experiments we were finding instead that this conclusion is, almost always, only approximately true. But in probability calculus is actually the very theory that cannot deny the possibility of all frequencies, since every frequency is compatible with all possible assessments of the probability of its occurrence: uncertainty is in fact inside the theory, while in other sciences it depends on the faulty agreement between theory and empirical facts.

4 "Small" probabilities and "falsification"

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.

Ruling out the possibility of those cases which seem "exceptional", in no way "improves" the probabilistic explanation (by somehow making it simpler, or "more scientific"): on the contrary, it is a way of denying it.

We could say that a "law" which is based on a probabilistic interpretation does not play the role of explaining why a fact that we forecast will occur, but rather of explaining why we forecast that a fact will occur (it is not a pun, but the essence of the argument!): when we press "at random" 18 keys on a typewriter, we are not able to explain why the statement "to be or not to be" has not been written (no "physical" cause ...); on the contrary, 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* as that piece of "Hamlet" of being typed.

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 their "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. Acceptance of the probabilistic interpretation of a class of phenomena entails that what we state must not be regarded as *necessary*, but instead regarded as only *probable* (even if "highly probable"). So it is *not* a question of asserting the existence of *exact laws which are obeyed only approximately:* one must consider as "natural" (*i.e.*, not impossible) also departures from any rigid law, occurrence of fluctuations, the effect of possible discontinuities,... and so on. In fact the history of science offers many examples of important discov-

eries that had their origin in the perception of someone who saw a (possibly small) unexpected variation in his data. 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). For example, if after 200 draws from a box containing 90 balls numbered from 1 to 90 the outcome 27 did not occur (an event which is not so rare, as the gamblers of Italian lotto know ...), is there someone willing to modify the "probabilistic law" stating that the probability of the outcome 27 is 1/90?

An interesting example is dealt with in a paper by Jaynes³, who claims that what happens in the real world is on the level of ontology (*i.e.*, it has an "objective" existence), while "what we can predict depends on our state of knowledge and is necessarily on the level of epistemology". He consider the status of the second law in biology, which exhibits some mysteries (to one who thinks that probabilities are physically real things): he makes a detailed study of a seeming contradiction between animal muscle efficiency and the second law, arguing against the "tendency to disorder" acritically attached to the second law (Nature has no way of knowing what we consider "orderly").

Jaynes' paper contains not only a detailed discussion, but also detailed computations to show that a living cell (like a muscle) does not violate the second law, if the latter is suitably interpreted: in fact its behaviour arises from a deep interplay between the epistemological macrostate (*i.e.*, the variables like pressure, volume, magnetization, that an experimenter measures and which therefore represent our knowledge about the system) and the ontological microstate (the coordinates and momenta of individual atoms, which determine what the system will in fact do). The initial microstate is unknown because it is not being controlled by any of the macroscopic conditions. The essential conclusion of Jaynes is that the increase of entropy cannot be an ontological statement (*i.e.*, a deductively proved consequence of the laws of physics): instead of committing the error of supposing that a given physical system has one and only one "true" ontological entropy, we must recognize that we could have many different states of knowledge about it, leading to many different entropies.

5 Conditional probability

Another important aspect to be dealt with is the following: seemingly paradoxical situations may depend on the arbitrary identification of mutually exclusive conditional frequencies with a (unique) conditional probability function. The perspective under which the problem will be discussed is essentially based on a statement which is, in a sense, very obvious, namely: *one should be care-* ful when applying the rules of probability theory to observed frequencies (in general, and so in particular when the latter refer to mutually incompatible experiments).

Definition – Let \mathcal{E} and \mathcal{H} be two arbitrary families of events, with $\Omega \in \mathcal{E}$ and $\emptyset \notin \mathcal{H}$. A nonnegative function $P(\cdot|\cdot)$ is said a *weak conditional probability* on $\mathcal{E} \times \mathcal{H}$ if

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

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

(b) P(H|H) = 1 for any $H \in \mathcal{E} \cap \mathcal{H}$;

(c) given E, H, A such that $E \in \mathcal{E}, A \in \mathcal{E}, E \cap A \in \mathcal{E}, H \in \mathcal{H}$ and $E \cap H \in \mathcal{H}$, 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 *weak 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

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

We referred to an arbitrary set $C = \mathcal{E} \times \mathcal{H}$ of conditional events, with no underlying structure. The function P is called a (finitely additive) **conditional probability** when \mathcal{E} is an algebra and $\mathcal{H} \cup \{\emptyset\}$ is a subalgebra of \mathcal{E} .

The following results are well known (see 4,5,6):

(i) a conditional probability is coherent;

(*ii*) coherence of a function P on an arbitrary family C of conditional events entails that P is a weak conditional probability on C (so that the "axioms" (a), (b), (c) play the role of *necessary* conditions for coherence).

(*iii*) if C is a given family of conditional events and P a corresponding assessment, then there exists a (possibly not unique) coherent extension of P to an arbitrary family \mathcal{G} of conditional events, with $\mathcal{G} \supseteq \mathcal{C}$, if and only if P is coherent on \mathcal{C} .

It follows easily (just referring to the minimal algebra containing C) that:

(*iv*) a function P on C is coherent *if and only if* P is the restriction of a conditional probability.

This statement singles-out the *syntactic* part of the concept of coherence. So it is possible to assess your degrees of belief on the basis of the syntactic rules, *with no reference to the device of betting*.

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 (coherence!) must then be taken into account.

6 Misunderstandings in conditional events and probabilities

Some **fundamental** remarks are now in order. First of all, notice that *property* (c) of conditional probability is crucial in order to fully exploit its inferential meaning. In fact, what it is usually 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" Ω . 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 (see⁷, p.284) concerning, for a given conditional event H|E, the "information" represented by E: the main point is a distinction between "assumed" and "acquired" E. For example, in Bayesian inferential statistics, given any event H (seen as hypothesis), with prior probability P(H), and a set of events E_1, \ldots, E_n representing the possible statistical observations, with likelihoods $P(E_1|H), ..., P(E_n|H)$, all posterior probabilities $P(H|E_1), ..., P(H|E_n)$ 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, $P(H|E_k)$ is chosen – among the prearranged posteriors – as the updated probability of H: 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 $P(H|E_k) = p$, to regard a conditional event H|E as a whole and interpret p as – look at the position of the brackets! – "the probability of (H given E)" and not as "(the probability of H), given E". On the other hand, the latter interpretation is unsustainable, since it would literally mean "if E occurs, then the probability of H is p", which is actually a form of a logical deduction leading to absurd conclusions.

For example, consider a set of five balls $\{1, 2, 3, 4, 5\}$ and the probability of the event H 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 H 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.

7 Probability and urns

Let us now discuss the following example (our discussion has a close connection with a famous statement – "Probability does not exist" – put at the very beginning of his book by B. de Finetti¹ (1974, Preface): in spite of its provocative and striking flavor, it is not difficult to encounter real and very simple accomplishments of that statement). 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**, since he does not acknowledge E as an event.

De Finetti's position is that in principle no distinction can be done, since probabilities exist only inasmuch as one creates them as useful substitutes for *a lack of information about something* (namely, for *uncertainty*): no "physical" or combinatorial interpretation is needed (though they are not forbidden, under suitable constraints). Given any set of events whatsoever (recall – see Sect.2 – that our concept of event identifies it with a *proposition*, and not necessarily with a subset of the so-called "sample space"), 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)
$$P(E) = \sum_{r=0}^{N} P(H_r) P(E|H_r)$$

In particular, we may assume (for example) that the given box has been chosen at random among N+1 boxes corresponding to all possible compositions: 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)
$$P(E|H_r) = \frac{r}{N},$$

we may conclude, by straightforward computations, that

$$P(E) = \frac{1}{2}$$

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 (even if someone could

prefer to speak of an *a priori* evaluation), 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).

8 Is frequency a coherent extension?

Observed frequencies (pertaining to different experiments) may not be necessarily identified with (and so used to compute) probabilities, and the previous discussion can be seen as an instance of the problem of *finding a coherent extension of some beforehand given probabilities* (see result *(iii)* in Sect.5!).

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 (a sort of "weighted average") of conditional *probabilities* can be – see eq.(1) – a *probability*, a convex combination of conditional *frequencies* is not a frequency (furthermore one 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 (for one or more events) even only one of the conditions (a), (b), (c), 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.

9 Probability in Quantum Mechanics

Let us now consider the following example. Take as "physical system" a ball drawn from a box \mathbf{U} of *unknown composition* containing 4 balls, each one being either white or black; as "final state" take

A = a ball drawn from the box **U** is white.

The given "experimental situation" is the following: we have built up the given box by taking a box \mathbf{K} of known composition with 3 balls (2 black and 1 white), and inserting in it one more ball (taken at random from a bag con-

taining an equal number of black and white balls) without looking its colour. To alter the experimental conditions we may consider as a "suitable device" *just looking at the colour of the inserted ball.* Then put

 S_1 = the inserted ball is white ,

 S_2 = the inserted ball is black .

Notice (cfr. the beginning of Sect.6) that A, S_1, S_2 correspond to *uncertain* events, and not just to given *facts*.

Now consider a large number of drawings with replacement from the given box (i.e., that with 4 balls) under the situation S_i (i = 1, 2): evaluating the (conditional) probability of A by the observed frequency gives, "almost certainly",

$$P(A|S_1) = \frac{1}{2}$$
, $P(A|S_2) = \frac{1}{4}$.

On the other hand

$$P(S_1) = P(S_2) = \frac{1}{2},$$

so that we should have, by (\star) ,

$$P(A) = P(S_1)P(A|S_1) + P(S_2)P(A|S_2) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{4} = \frac{3}{8}.$$

Can the above value be really equal to P(A)? Notice that the assessment of P(A) requires referring to a single experiment and that the colour of the inserted ball is not known. But pretending an "experimental measure" of P(A), necessarily obtained by drawing with replacement a large number of times from the given box **U**, leads to a "long run" observed frequency that is almost certainly equal either to 1/2 or to 1/4, and **not** to 3/8.

So a strict frequentist evaluation makes the latter value meaningless, since the box contains either 1 or 2 white balls out of 4. In conclusion, the lack of validity of formula

(4)
$$P(A) = P(S_1)P(A|S_1) + P(S_2)P(A|S_2),$$

depends on the identification of the probabilities $P(A|S_1)$ and $P(A|S_2)$ with observed frequencies (corresponding to different and incompatible experiments), each one referring to a different box obtained by inserting a given ball (whose colour, white or black, is known) in the box **K** initially containing three balls.

Nevertheless these experiments are not (so to say) "mentally" incompatible if we argue in terms of the general interpretation of probability (for example, $P(A|S_1)$ is the degree of belief in A under the assumption – not necessarily an observation, but just an assumed state of information – " S_1 is true"): then, for a coherent evaluation of P(A) we must necessarily rely only on the above value obtained by resorting to the second member of (4), even if such probability ... "does not exist" (in the sense, already discussed at the beginning of Sect.7, that it does not express any sort of "physical property" of the given box).

Also 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 Feynman⁸(1951), 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 Scozzafava⁹(1991).

A source emits "identically prepared" particles toward a screen with two narrow openings, denoted S_1 and S_2 . Behind the screen there is a film which registers the relative frequency of particles hitting a small given region A of the film. Measurements are performed in three different physical situations: both slits open, only slit S_1 open, only slit S_2 open.

Let us introduce, for a given particle, the following event, denoted (by abusing notation) by the same symbol of the corresponding physical device:

A = the particle reaches the region A,

and, for i = 1, 2, the following two events:

 S_i = the particle goes through slit S_i .

Moreover, since all the particles are identically prepared, we shall omit the further symbol H (referring to preparation) in all conditioning events.

The aforementioned experimentally measured frequencies are usually identified, respectively, with the three probabilities P(A), $P(A|S_1)$ and $P(A|S_2)$. Repeated experiments can be performed letting a particle start from the source, and then measuring its final position on the film, to determine whether it is in the region A or not; moreover we could "measure" $P(A|S_1)$ or $P(A|S_2)$ letting be put in function an experimental device allowing the particle going to hit the region A only through the slit S_1 or only through the slit S_2 . The latter corresponding frequencies (of going through the relevant slit) are also identified with the probabilities $P(S_1)$ and $P(S_2)$. Now, irrespective of whether the device has been activated or not, and of what was the issue in case of activation, we may obviously write eq.(4), which is (as it has been exhaustively discussed) an elementary property of probability, easy consequence of axioms (a) and (c). Instead physical experiments give an inequality between left and right hand side of (4), **but this circumstance cannot be used to** "falsify" anything, since it refers in fact only to observed frequencies. Those who consider this phenomenon as strictly peculiar to quantum physics, as $Jordan^{10}(1926)$, try to relate it to the well-known anomalies encountered in situations where, depending on the relevant specific aspects, either the so-called "particle interpretation" or the so-called "wave interpretation" is more appropriate. For example, in this experiment an *interference* could be expected from a wave motion going simultaneously through both slits. However a classical wave motion cannot explain the phenomenon, since it is possible to let only one particle at a time reach the apparatus, and still the measurements show an inequality holding in place of (4).

In more general terms, we have essentially a physical system that may reach a final state under a given experimental situation, which in turn can be altered by activating a suitable experimental device. (In the case of the two-slit experiment, the particle can be regarded as a system whose final state is its reaching the region A of the film under the given experimental situation corresponding to both slits open; it can be altered by activating a device letting only slit S_1 or only slit S_2 open).

The above situation has an elementary analogue in the example discussed at the beginning of this Section, where there is no quantum effect to which the lack of validity of formula (4) could be ascribed: notice that we adopted the same notation and symbols, to help the identification of similar corresponding events. Clearly, in the case of both slits open, if we look at P(A) by referring to a single particle, we can argue as in the aforementioned example. On the other hand, referring to repeated experiments, much care is needed when dealing with frequencies and (conditional) 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. Not to mention that this choice – as it has already been discussed in the beginning of Sect.6 – appears as a very particular case of conditioning, whose meaning is just that of *changing the sample space*.

10 References

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

2. R. Scozzafava, "A merged approach to stochastics in engineering curricula", *European Journal of Engineering Education*, 15 (3), 241-248 (1990).

3. E.T. Jaynes, "Clearing up mysteries – The original goal", Maximum

Entropy and Bayesian Methods (J.Skilling, Ed.), Kluwer Academic Publishers, 1–27 (1989).

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

5. S. Holzer, "On coherence and conditional prevision", *Boll. Un. Mat. Ital.*, (6) 4, 441-460 (1985).

6. G. Coletti and R. Scozzafava, "Characterization of coherent conditional probabilities as a tool for their assessment and extension", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4, 103-127 (1996).

7. G. Coletti, D. Dubois, and R. Scozzafava, *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, Plenum Press, New York (1995).

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

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

10. P. Jordan, "Über eine neue Begründung der Quantenmechanik", Zeitschrift für Physik, 40, 809-838 (1926).