Subjective Conditional Probability and Coherence Principles for Handling Partial Information

Romano Scozzafava
Dipartimento Metodi e Modelli Matematici
Universita' "La Sapienza" - Roma, Italy

1 Introduction

In the most widely used approach to the theory of probability, based on a measure-theoretic framework, the set of all possible envisaged situations (the so-called sample space) needs to be *endowed with a given "algebraic" structure* (such as a boolean ring, a σ-algebra, etc.), on which an *overall probability assignment* is defined. Not to mention that this setting usually entails the further consequence of assuming (without any sound reason, apart from mathematical convenience) *countable additivity* of the probability. On the other hand, for many real world situations it is suitable and very significant not assuming any specific structure for the set where probability is assessed and not making such assessment on the whole set of possible situations.

In this respect, the theory of probability as proposed by de Finetti (based on the concept of *coherence*) is particularly suitable: it allows to assess your (coherent) probability for as many or as few events as you feel initially able and interested, possibly going on to further events, and this has many important theoretical and applied consequences, making simpler and more effective the "operational" aspects. In fact checking of coherence amounts to the study of the compatibility of some linear systems, whose unknowns are the probabilities of the atoms generated by the given events.

The link between probability measures and coherent functions is the following: *coherence* (which can be defined in terms either of a *penalty criterion* or of a *betting scheme*: for the former see de Finetti (1970) and Gilio (1990), the latter is discussed in Sect.4) of a function P on an arbitrary family $\mathcal{E}$ of events is equivalent to the existence of a probability space $(\Omega, \mathcal{A}, P')$, with a finitely additive P' and the Boolean algebra $\mathcal{A} \supseteq \mathcal{E}$, such that the given function P is the restriction of P' to $\mathcal{E}$.

But the main aim of this paper is to emphasize the *semantic* aspects of this approach rather than a detailed discussion of the aforementioned *syntactical* aspects (for these, we refer to the paper by G.Coletti in this same issue, which should be read in conjunction with this one).

In fact, a clear-cut distinction between the *meaning* of probability and the various multifacet *methods of assessment* is essential. With respect to the meaning, probability can be regarded as a measure of the *degree of belief* hold by the subject that is making the assessment (this is the essential reason why it is called *subjective* probability).

2 Probability: meaning and assessment

The most "popular" and well known methods of assessment are based on the *combinatorial* approach and on the *observed frequency.* These two methods are dealt with in any text book on probability (even if under the misleading attribute of "definitions": see de Finetti (1972, 1976) and Scozzafava (1990a) for a relevant discussion), and so there is no need to spend much time on them.

Yet note that they essentially suggest to *take into account only the most schematic data or information, and in the most schematic manner*, which is not necessarily bad, but not necessarily good either. Nevertheless, these most "popular" and well known approaches to probability must not be doomed, but may be taken as useful methods of assessment: they are subjective as well, since it is up to the given *subject* to judge, for example, the "symmetry" in the combinatorial approach or the existence of "similar" conditions for the different trials in the frequentist approach.

A more general conceptual framework can be based on inductive reasoning, which is a fundamental tool also for the scientific knowledge: an instance is the possibility of introducing a well founded process of measuring the expectation of "future" events on the basis of observed "past" events (usually, statistical data). In the various real-life situations in which uncertainty is present, gathering and interpreting statistical data leads in general to a decrease of uncertainty with respect to the initial situation: the measurement of this uncertainty can be quantitatively carried out by the same tool used for the measurement of the uncertainty associated with classical random phenomena such as coin tossing or dice throwing, i.e. through the concept of *probability.* The outlined process may be called, in a vague but expressive way, "learning from experience", and its precise and rigorous formulation can in fact be realized by *conditional* probability.

Statistics and probability proceed so at the same pace, the former providing techniques for the formalization and the synthesis of data, the latter interpreting them through "conclusions" which in general *are not certain* (as those of ordinary, i.e. deductive, logic), but only more or less probable.

In order to fully grasp the richness of this merged approach, an overcoming of barriers created by prevailing opinions (that rely upon a *"combinatorial" assessment*, assuming equal probability of all possible cases, or upon the assessment of the *probability of an event* through the *frequency* relative to other events that are considered, in a sense, "equal" to that of interest) is needed. In other words, it is essential to give up any artful limitation of probability to particular events (not even clearly definable), that often unnecessarily restricts its domain of applicability. There is instead the need to ascribe to probability a more general meaning, which after all should be a sensible way to cope with real situations: a concrete and not stereotyped approach should in fact start from the subjective and intuitive "real life" meaning of probability as *degree of belief* in the occurrence of an event.

This can be easily done through a natural condition of *coherence*, so leading to the conclusion that subjective probability satisfies the usual and classical properties: it is a function whose range is *between zero and one* (these two extreme values being taken, in particular, by the *impossible* and *certain* event respectively) and which is *additive* for mutually exclusive events. These properties constitute the starting point in the axiomatic approach: in conclusion, *the subjective view can only enlarge and never restrict the practical purport of probability theory.* For a deeper discussion, see also Scozzafava (1989, 1991).

3 Conditional events as uncertain statements in Artificial Intelligence

How can the probabilistic approach to Artificial Intelligence be handled according to the theory sketched above? In this paper we deal with and emphasize only some aspects: other relevant papers are Gilio and Scozzafava (1988), Coletti, Gilio and Scozzafava (1991, 1993), Gilio and Spezzaferri (1992), Coletti (1993, 1994), Coletti and Scozzafava (1993).

First of all, a formulation of uncertain statements in terms of *conditional* events is needed. We start by introducing the simpler concept of *event*, which can be singled-out (in our general framework) by a (nonambiguous) *proposition* E, that is a statement that can be either *true* or *false* (in particular, there is no need to distinguish between hypotheses and evidence). In general *it is not known* whether E is true or not: we are *uncertain* on E.

Some examples follow:

(1) a proposition E describing the so-called *"favorable"* cases to a possible outcome: so *outcome* and *proposition* can be mutually identified (typical situation is that of the combinatorial approach);

(2) given a sequence of trials performed "under similar conditions", a proposition E describing the possible result of each trial: so *observed result* and *proposition* can be mutually identified (typical situation is that of the frequentist approach);

(3) anything else (if sensible...).

Two important particular cases (with no uncertainty) are: the *certain* event Ω (where certain means that *it is known* that the relevant proposition Ω is true) and the *impossible* event ϕ (when *it is known* that the relevant proposition ϕ is false).

In all other cases, the actual "value" (true or false) of the event E is not known. Notice that, depending on the relevant "information" (for instance, statistical data or evidence), an uncertain event E may become true or false, so reducing to Ω or ϕ.

In general, *it is not enough directing attention just toward the event E in order to assess "convincingly" its probability:* it is also essential taking into account other events which may possibly contribute in determining the "information" on the event E.

Then the fundamental tool will be *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, acquired or assumed).

This requires the introduction of *conditional* events, which correspond to a 3-valued logic (see Sect.4): so it is necessary to define appropriate logical relations and operations, extending the usual ones between standard events. There are many *pros* and *contras* concerning the "right" choice among different possible definitions: usually they should depend on each specific context and application. For a deepening of these aspects, see Gilio and Scozzafava (1994), Goodman, Nguyen and Walker (1991). Anyway, real world situations make very significant assuming an "open" framework and not a specific algebraic structure for the family of conditional events on which probability is assessed.

The following example, due to Schay (1968), helps in throwing light on the aforementioned need of an open framework and of a careful interpretation of the concept of event as a *proposition*.

Given an election with only three candidates A, B, C, denote by the same symbols also the events corresponding to either one of the candidates winning, so that $A \cup B \cup C = \Omega$, the certain event. Now, suppose that C withdraws and that then all his votes will go to B: according to Schay, this situation involves probabilities for which the product rule

$$P(B \cap H) = P(B|H)P(H), \tag{1}$$

where $H = A \cup B$, *does not hold*. We are going to discuss (and challenge) this conclusion: it is based on the presumption that the new framework corresponding to the withdrawing of C is $A \cup B$.

Actually a careful singling-out of the "correct" events is needed: in our "open" framework this can be done also outside the initial "space" $\{A, B, C\}$, considering a suitable *proposition* (expressing the new information), which is not $A \cup B$, but the event

$E = C$ *withdraws and all his votes go to* B.

Notice that $E \subset A \cup B$ *but not conversely*: in other words, either A or B can win even if C does not withdraw!

Schay argues as follows: since one has $P(H) = 2/3$ and

$$P(B|H) = \frac{2}{3}, \tag{2}$$

from $B \cap H = B \cap (A \cup B) = B$ it follows, for the left member of (1), the value $P(B) = 1/3$, while the right member of the product rule is equal to $(2/3)(2/3) = 4/9$.

On the contrary, a careful singling-out of the conditioning event entails that we must replace (2) by

$$P(B|E) = \frac{2}{3},$$

with $E =$"*C withdraws and all his votes go to B*". Moreover, as shown in Scozzafava (1993), the only coherent assessment of $P(B|H)$, given the values $P(B|E) = 2/3$, $P(H) = 2/3$ and $P(B) = 1/3$, is $P(B|H) = 1/2$: then the product rule (1) holds.

Therefore, Schay's conclusion that *"it may along these lines be possible to incorporate the probabilities of quantum mechanics in our theory"* must be challenged. It has been discussed elsewhere (Scozzafava, 1992) how certain paradoxes, concerning probabilities that do not satisfy the product rule and arising in the statistical description of quantum theory, depend on the fact that observed frequencies, relative to *different* experiments, are arbitrarily *identified* with the values of a conditional probability on the *same* given space.

Moreover, there are many multifacet controversial aspects, including the need to avoid misunderstandings when the conditioning event is interpreted as a fit representation of a given information. For example, a careful distinction between *assumed* and *acquired* information is essential, i.e. an important issue concerns the need of interpreting the conditional probability $p = P(A|B)$ as

the probability of (A *given* B)

rather than as

(the probability of A) *given* B.

The latter interpretation is unsustainable, since it would literally mean *"if B occurs, then p is the probability of A"*, which is actually a form of logical deduction leading to absurd conclusions. Consider in fact a set of five balls $\{1, 2, 3, 4, 5\}$ and the probability (which equals 2/5) that a number drawn from it at random is even: this probability could instead be assessed equal to 1/3, since this is the value of the sought probability *conditionally on the occurrence of each one of the events*

$$E_1 = \{1, 2, 3\} \text{ or } E_2 = \{3, 4, 5\},$$

and one (possibly both) of them will *certainly* occur.

4 Interpretation of conditional events and conditional probability in terms of a betting scheme

A conditional event $E|H$, with $H \neq \phi$, is looked upon as a 3-valued entity taking one of the three values 1, 0, p (denoting by p any number between 0 and 1), according to whether it is true, respectively, either the event EH, or E^cH, or H^c.

If an amount p is paid to bet on $E|H$, we get, *when H turns out to be 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, which can be interpreted as the *conditional probability* $P(E|H)$: see Sect.5.

In particular, the (unconditional) event E, that can take only two values (1 or 0), can be looked on as $E|\Omega$.

5 Probability assessment and coherence

Let $\mathcal{C}$ be an arbitrary family of conditional events and P a real function defined on $\mathcal{C}$. Given *any* finite subfamily

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

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

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

(i.e, the gain corresponding to a combination of n bets of amounts $p_1\lambda_1, ..., p_n\lambda_n$ on $E_1|H_1, ..., E_n|H_n$, with arbitrary real stakes $\lambda_1, ..., \lambda_n$). 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 $P : \mathcal{C} \to R$ is coherent if, for each assessment $\mathcal{P} = (p_1, ..., p_n)$ on a finite family $\mathcal{F} \subseteq \mathcal{C}$, with $p_i = P(E_i|H_i)$, and for every choice of $\lambda_1, ..., \lambda_n \in R$, the possible values of the corresponding gain $G_{|H_0}$ are neither all positive nor all negative.*

Given the atoms $A_r (r = 1, 2, ..., m)$ generated by the $2n$ events $E_1, ..., E_n$, $H_1, ..., H_n$, the possible values of $G_{|H_0}$ are those corresponding to the partition of Ω into the atoms.

Notice that this result 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.* The point of defining probability in terms of hypothetical bets is to give it an unmistakable, concrete or operational meaning, but you may assess it however you like (obviously, you are not allowed to violate the relevant *syntactical* rules!).

It has been essentially proved by B. de Finetti that if the assessment P is coherent, then P can be looked on as the restriction on $\mathcal{C}$ of a (finitely additive) *conditional probability* given on $\mathcal{E} \times \mathcal{H}$, where $\mathcal{E}$ is a field and $\mathcal{H}$ an additive class.

The latter result and the following extension theorem, also essentially due to de Finetti (see de Finetti, 1949), constitute two milestones in the approach to probability through coherence.

Theorem *Given an assessment P on a class $\mathcal{C}$ of conditional events and an arbitrary class $\mathcal{K} \supseteq \mathcal{C}$, then there exists a (possibly not unique) coherent extension of P to $\mathcal{K}$ if and only if P is coherent on $\mathcal{C}$. In particular, if $\mathcal{K} = \mathcal{C} \cup \{E|H\}$ and $P(E|H) = p$, coherent assessments of the conditional probability p are those of a suitable closed interval $[p', p"] \subseteq [0,1]$, with $p' \leq p"$.*

So it is possible to assess P only on an arbitrary set $\mathcal{C}$ of conditional events of interest, with no underlying structure, and then to extend the assessment, preserving coherence, by a step-by-step assignment to further events. Moreover, since the conditional probability $P(E|H)$ is directly introduced as a function on the set of conditional events, bound to satisfy only the requirement of coherence, it can be assessed and makes sense for *any pair* of events E, H, with $H \neq \phi$, and there is *no need of assuming positive probability* for the conditioning event, as in the usual approach, where $P(E|H)$ is instead introduced by definition as the ratio between the two (unconditional) probabilities $P(E \cap H)$ and $P(H)$. In the "continuous" case the usual approach is based on the classical Radon-Nikodym framework, *which requires the knowledge of the whole conditioning distribution* (a situation which is clearly unsound, especially from a Bayesian inferential point of view), while we need refer just to the given conditional event: for a thorough critical comparison between de Finetti's and Kolmogorov's approaches, see Scozzafava (1990b).

The possibility of dealing also with zero probabilities is a very crucial feature, even in the case of a finite family of events: in fact, ignoring the possible existence of null events (which amounts to a stronger form of coherence) *drastically restricts the class of admissible probability assessments and the possibility of unboundedly extending a coherent conditional probability:* cf. the paper by G.Coletti in this same issue.

6 A sketch of conditional subjective probability looked on as membership function in fuzzy theory

In general, in the literature on fuzzy sets it is challenged the possibility of interpreting a statement such as $E =$ "Mary is young" as an event and the values of the corresponding membership function as probabilities. In fact E is a vague statement, and vagueness is looked on as an uncertainty about intended meaning and not about facts, i.e. a sort of "linguistic" uncertainty.

The arguments brought forward to distinguish grades of membership from probabilities obviously refer to the usual restrictive interpretations of event and probability: we will deepen in a forthcoming paper how our probabilistic framework allows to overcome this (putative) distinction in terms of conditional events and conditional probability.

In this paper we give only a sketch of our views by referring to the above classical example of fuzzy statement: where does it come from and what is its "operational" meaning?

From a pragmatic point of view, it is natural to think to some available information about possible values of Mary's age, which allows to refer it to a (subjective) membership function of the fuzzy set of "young" people. For example, for values of the age less than 25 the membership function may be put equal to 1, while it is put equal to 0 for values greater than 40; then it is taken as decreasing from 1 to 0 in the interval from 25 to 40.

One of the putative merits of the fuzzy approach is that, given the range of values from 0 to 1, there is no restriction for the assignment of the membership function, in contrast to probability that obeys certain rules such as, for example, the axiom of additivity: it follows that, when an expert assigns a subjective probability of (say) 0.2 to the statement that Mary's age is between 35 and 36, he inescapably assigns a degree of belief of 0.8 to the contrary, and he may not have for the latter fact any justification apart from the consistency argument represented by the additivity rule.

In our probabilistic framework the way-out is indeed very simple. Notice that the above choice of the membership function implies that women whose age is less than 25 are "young", while those with an age greater than 40 are not. So the real problem is that we are uncertain on being or not "young" those women having an age between 25 and 40: then our interest is in fact directed toward *conditional events* such as $E|H$, with

$$E = \textit{One claims that Mary is young}$$

$$H_n = \textit{the age of Mary is between } n \textit{ and } n+1,$$

where n varies over the integers from 25 to 39. It follows that an expert may assign a subjective probability $P(E|H_{35})$ equal to 0.2 without any need to assign a degree of belief of 0.8 to the event E under the assumption H_{35}^c (i.e., the age of Mary is not between 35 and 36), since an additivity rule with respect to the *conditioning* events *does not hold.* In other words, it seems sensible to identify the values of the membership function with suitable conditional probabilities: in particular, putting

$$H_0 = \textit{Mary's age is greater than 40,}$$

$$H_1 = \textit{Mary's age is less than 25 ,}$$

then $E \cap H_0 = \phi$ and $H_1 \subseteq E$, so that $P(E|H_0) = 0$ and $P(E|H_1) = 1$.

References

G. Coletti (1993), Numerical and qualitative judgments in probabilistic expert systems, in: *Proceedings International Workshop SIS on "Probabilistic Methods in Expert Systems"* (Ed. R.Scozzafava), Roma, Italy, 37-55.

G. Coletti (1994), Coherent Numerical and Ordinal Probabilistic Assessments, *IEEE Transactions on Systems, Man, and Cybernetics,* **24**, n.12, 1747-1754.

G. Coletti, A. Gilio, and R. Scozzafava (1991), Conditional events with vague information in expert systems, in: *Lecture Notes in Computer Sciences* (Eds. B.Bouchon-Meunier, R.R.Yager, L.A.Zadeh), Springer- Verlag, n.521, 106-114.

G. Coletti, A. Gilio, and R. Scozzafava (1993), Comparative probability for conditional events: a new look through coherence, *Theory and Decision*, **35**, pp.237-258.

G. Coletti and R. Scozzafava (1993), A Coherent Qualitative Bayes' Theorem and its Application in Artificial Intelligence, in: *Proc. of "2nd International Symp. on Uncertainty Modeling and Analysis"* (Ed. B.M.Ayyub), Maryland, U.S.A., IEEE Computer Society Press, Los Alamitos, California, 40-44.

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

B. De Finetti (1970), Teoria della probabilita' (Torino, Einaudi); English transl. (1974): *Theory of Probability,* Voll.1 & 2 (Chichester, Wiley).

B. De Finetti (1972), Subjective or objective probability: is the dispute undecidable?, *Symposia Mathematica,* 9, 21-36.

B. De Finetti (1976), Probability: beware of falsifications!, *Scientia,* 111, pp. 282-303.

A. Gilio (1990), Criterio di penalizzazione e condizioni di coerenza nella valutazione soggettiva della probabilita', *Boll. Un. Mat. Ital.* (7), **4-B**, 645-660.

A. Gilio and R. Scozzafava (1988), Le probabilita‘ condizionate coerenti nei sistemi esperti, in: *Atti Giornate A.I.R.O. su "Ricerca Operativa e Intelligenza Artificiale"* (Pisa, IBM), 317-330.

A. Gilio and R. Scozzafava (1994), Conditional events in probability assessment and revision, *IEEE Transactions on Systems, Man, and Cybernetics,* **24**, n.12, 1741-1746.

A. Gilio and F. Spezzaferri (1992), Knowledge integration for conditional probability assessments, In: *Uncertainty in Artificial Intelligence* (Eds. D.Dubois, M.P.Wellman, B.D'Ambrosio, and P.Smets), San Mateo: Morgan Kaufmann, 98-103.

I.R. Goodman, H.T. Nguyen, and E.A. Walker (1991), *Conditional Inference and Logic for Intelligent Systems: a theory of measure-free conditioning* (Amsterdam, North-Holland).

G. Schay (1968), An Algebra of Conditional Events, *Journal of Mathematical Analysis and Applications,* **24**, 334-344.

R. Scozzafava (1989), *La probabilita' soggettiva e le sue applicazioni* (Milano, Masson: 3 Ed. 1993).

R. Scozzafava (1990a), A merged approach to stochastics in engineering curricula, *European J. of Engineering Education,* **15**, n.3, 243-250.

R. Scozzafava (1990b), Probabilita‘ condizionate: de Finetti o Kolmogorov?, In: *Scritti in omaggio a L. Daboni,* Trieste: LINT. 223-237.

R. Scozzafava (1991), Probability assessment, in: *Proc. of International School of Math. "G.Stampacchia" on "Probability and Bayesian Statistics in Medicine and Biology",* Erice, Italy (Eds. I.Barrai, G.Coletti, M.Di Bacco), CNR Applied Math. Monographs n.4, 244-255.

R. Scozzafava (1992), A classical analogue of the 2-slit model of quantum probability (Invited paper at the 20th Europ. Meeting of Statisticians, Bath, U.K., 1992), *Pure Mathematics and Applications,* Series C, 2, 223-235.

R. Scozzafava (1993), How to solve some critical examples by a proper use of coherent probability, in: *Uncertainty in Intelligent Systems* (Eds. B.Bouchon-Meunier, E.Valverde, R.R.Yager), Amsterdam: Elsevier, 121-132.