Ratio Math. n. 7 - 1994
75 - 82

A REAL LIFE APPROACH TO THE TEACHING OF PROBABILITY (*)

Romano Scozzafava
Dipartimento Metodi e Modelli Matematici
Università "La Sapienza" - Roma (Italy)

What are the challenges and perspectives for probability and statistics education, apart from the need of providing a professional tool for working on practical problems? Since classical mathematics is still a source of intriguing and stimulating ideas, it seems inadequate to subscribe to the conception that the role of probability and statistics is that of making pupils acquainted with the existence of new fundamental topics besides classical mathematical theories. In other words, we maintain that the main role of this teaching is essentially that of introducing a *new way of thinking* (i.e. inductive, to handle random situations) more than that of acquiring some interesting and less usual mathematical notions or that of emphasizing a utilitarian point of view. The conceptual framework of inductive reasoning, which is fundamental for the scientific knowledge, lies in the possibility 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 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." Statistics and probability proceed so at the same pace, the former providing techniques for the normalization 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

---

(*) Contributed paper at ICME-7, Québec, 1992.

more or less probable. In order to fully grasp the richness of this merged approach to probability and statistics, an overcoming of barriers created by prevailing opinions is needed: for example, many elementary approaches to these subjects rely essentially upon a "*combinatorial*" *assessment of probability* (assuming equal probability of all possible cases) and upon the possibility of introducing the *probability of an event* through the *frequency observed in the past* for other events that are considered, in a sense, "equal" to that of interest. Yet it is not generally underlined that the choice of these events (as the choice, in the combinatorial approach, of the outcomes for which equal probability is assumed) is necessarily *subjective*.

For example, an insurance company that needs to evaluate the probability of dying within the year of a given person can base its assessment on data referring to individuals of the same town (or region, or district) as the given person, or of the same age, or sex, or civil status, or of the same profession, or income, or having an analogous bodily constitution (height, weight, etc.), and so on, *grouping in many different ways some or all of the preceeding characteristics, or possibly others,* and to each of these (subjective) choices there corresponds in general a *different frequency*.

In other words, it is essential to give up any artful limitation to particular events (not even clearly definable) and try to ascribe to probability a more general meaning, which after all should be a sensible way to cope with real situations: in fact a concrete and not stereotyped teaching should start from the subjective and intuitive "real life" meaning of probability as *degree of belief* in the occurrence of an event. Last but not least, *from a didactic point of view* the introduction, through a natural condition of *coherence*, of *subjective probability* is very simple and quick. It is in fact easy to show (as we are going to do in what follows) 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: they are well-known for the two most "popular" cases, i.e. when the event of interest is embedded in a set of possible outcomes which are considered, in some (necessarily subjective) sense, "symmetric", or when the event is seen as a further repetition, under "similar" (subjective judgement!) conditions, of a given phenomenon.

An *event* E is any unambiguous proposition that can only assume the two "values" TRUE or FALSE (denoted by 1 or 0, if E is regarded as a simple *random variable*). The lack of information on the actual value of E paves the way to the introduction, as an ersatz, of the concept of probability.

A value $p = P(E)$ (or pS, for arbitrary $S \neq 0$) is regarded as an amount to be paid to bet on E, with the proviso of winning a unit amount (or an amount S) of money if E occurs and nothing if E does not occur (in other words, p is the amount to be paid to get an amount *equal to the value assumed by* E): *coherence* is defined by the requirement that *the choice of* p *would not make the player a sure loser or winner*. If E is different from Ω and $\varnothing$ (*certain* and *impossible* event), the two possible "gains" are

$G_1 = (-p + 1)S$ (*if* E *occurs*),

$G_2 = -pS$ (*if* E *does not occur*),

and so, since coherence requires that they must not be both negative or both positive, p must satisfy the inequality $-p(1 - p)S^2 < 0$, which is, for any $S \neq 0$, the same as

$$0 < p < 1.$$

When $E = \Omega$ or $E = \varnothing$ there is *no uncertainty* on the outcome of the corresponding bet: the *only* (certain!) value of the gain is

$$G(\Omega) = (-p + 1)S \quad \text{or} \quad G(\varnothing) = -pS$$

respectively, and so coherence requires that the gain is *equal to zero*, which gives $p = 1$ for $E = \Omega$ and $p = 0$ for $E = \varnothing$. In conclusion, if the subjective probability of E (our *degree of belief* on E) is defined as *an amount* $p = P(E)$ *that makes coherent a bet on* E, then

$$(1) \qquad 0 \leq P(E) \leq 1,$$

with, in particular,

$$(2) \qquad P(\Omega) = 1, \quad P(\emptyset) = 0.$$

Now, given a finite *partition* of $\Omega = E_1 \vee E_2 \vee \ldots \vee E_n$ and n simultaneous bets, let $P(E_i)$ be the amount paid for a coherent bet on E_i (with $i = 1, 2, \ldots, n$): clearly, these n bets can be regarded as a single bet on Ω with amount $P(E_1) + P(E_2) + \ldots + P(E_n)$, and so the first of (2) implies

$$(3) \qquad P(E_1) + P(E_2) + \ldots + P(E_n) = 1.$$

So the usual "axioms" (1), (2), (3) of probability are easily obtained in a very simple way, as necessary and sufficient conditions for coherence. Notice that this result is based on *hypothetical* bets: the force of the argument *does not depend on whether or not one actually intends to bet*, since 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 "combinatorial" and "frequentist" methods of evaluation of probabilities can be easily embedded into the general concept of subjective probability. In fact, given n possible outcomes represented by the events $E_1, E_2, \ldots, E_n$ of a partition of Ω, and an event E which is a union of r among the E_i's, the classical evaluation $P(E) = r/n$ follows easily from (3) (and from its extension to the case in which the events are still mutually exclusive but not necessarily constituting a partition) *through the subjective opinion that a symmetry exists and that it implies equality of probabilities*, namely $P(E_i) = 1/n$. Moreover, for a sequence $A_1, A_2, \ldots, A_{n+m}$ of events ("trials" of a given phenomenon), assume that the outcome of the first m is known (i. e., the corresponding "past" *frequency* X is, say, k/m) and consider the "future" frequency

$$y = \frac{A_{m+1} + A_{m+2} + \ldots + A_{m+n}}{n}.$$

The *prevision* of a random variable like Y can be interpreted *similarly to the probability of an event*: $\mathscr{P}(Y)$ is the amount that must be paid (in n simultaneous coherent bets on the events $A_{m+1}, A_{m+2}, \ldots, A_{m+n}$ of amounts $P(A_i)S$, where $S = 1/n$) to get an amount *equal to the value that will assume* Y, i.e.

$$\mathscr{P}(Y) = \frac{P(A_{m+1}) + P(A_{m+2}) + \ldots + P(A_{m+n})}{n}.$$

If the above events are judged *equally probable* (subjective opinion!) we get, denoting by p this common probability

$$\mathscr{P}(Y) = p. \qquad (4)$$

Assuming (subjective opinion!) that *the probability distribution of the "future" frequency* Y *is equal to that of the "past" frequency* X, whose value is known and equal to k/m (so that $\mathscr{P}(X) = k/m$), from (4) it follows that $p = k/m$, i.e. the "frequentist" evaluation of $P(A_i)$ for $i \geq m+1$.

It is important to point out that this approach puts in the right perspective all the subjective aspects hidden in the so-called "objectivistic theories". Probability cannot be simply seen as a "physical" property of an event and may be considered also for "unique" situations, going beyond the two particular cases corresponding to "*equally probable outcomes*" and "*events that can be repeated under similar conditions*".

Let us now face the problem of choosing, for each event E, a suitable value of its probability p *among all coherent evaluations*, i.e. those satisfying (1), (2), (3): it is not enough directing our attention only toward the event E, but we need taking into account also *other events* which contribute in determining our information on E. To this end two fundamental tools are *conditional probability and Bayes' theorem.*

A bet on a *conditional event* $E|H$, with $H \neq \varnothing$, is a bet on E which is *called off* when H does not occur: so, if an amount p is paid to bet on $E|H$, we get, *when* H *turns out to be true*, an amount 1 if E is true and an amount 0 if E is false, and *we get back the amount* p *if* H *turns out to be false*. Then, putting $p = P(E|H)$, if such a bet is *coherent* the function

$$P_H(\cdot) = P(\cdot\,|H)$$

verifies (1), (2), (3) *for any given* $H \neq \varnothing$. In this case $P(E|H)$ is said *conditional probability* (of E given H). Recall that, *for unconditional bets*, p is the amount to be paid to get an amount *equal to the value assumed by* E (i.e. 1 or 0), or, more generally, the *prevision* $\mathscr{P}(X)$ of a random variable X is the amount to be paid to get an amount *equal to the value that will assume* X. It follows that $E|H$ can be regarded as a *three-valued event*, or else as a random variable X assuming one of the three values 1, 0, p, corresponding respectively *to won, lost, called off* bet, according to whether $H = E = 1$, or $H = 1$ and $E = 0$, or $H = 0$, that is

$$X = E\,H + p(1 - H).$$

Then the prevision of X must coincide with $p = P(E|H)$, i.e.

$$p = \mathscr{P}(X) = P(E \wedge H) + p - p\,P(H). \qquad (5)$$

From (5) it follows easily

$$P(E \wedge H) = P(H)\,P(E|H). \qquad (6)$$

Notice that this fundamental (and well-known) relationship involving conditional probability has been deduced, like (1), (2), (3), *assuming only coherence* (i.e. the obvious condition that no bet is sensible if one knows in advance to be a sure loser or winner!); moreover, *no restriction has been imposed on the kind of events involved* (such as equally probable, or "repeatable", or anything else).

By re-writing eq. (6) with H^c (the *opposite* of H) in place of H and summing (6) itself to the relation so obtained, we get, since $(E \wedge H) \vee (E \wedge H^c) = E$,

$$(7) \qquad P(E) = P(H)\,P(E|H) + P(H^c)\,P(E|H^c).$$

Interchanging the role of E and H in (6), we may eliminate the left-hand side from these two equations; then, assuming that $P(E) > 0$ and recalling (7), we obtain

$$(8) \qquad P(H|E) = \frac{P(H)\,P(E|H)}{P(H)\,P(E|H) + P(H^c)\,P(E|H^c)}.$$

This formula constitutes *Bayes' theorem*, that allows, given an event E (representing, for example, observed data), evaluating the conditional probability $P(H|E)$ of an other event H (often called *hypothesis*), which *initially* (i.e., before collecting data) had been assessed equal to $P(H)$. In other words, the true problem for any event H may be considered *not* simply that of evaluating $P(H)$, which may be possibly regarded just as a provisional assessment (a *prior*, in a more technical jargon), but that of evaluating $P(H|E)$ taking into account all the relevant information carried by some other event E.

So a "merging" of probability and statistics can be easily attained without being entangled with involved or complicated arguments. Thanks to the conditional probability $P(H|E)$ it is possible to give a different *probability evaluation* of H for each different "state of information" expressed by E (usually corresponding to *statistical* data): almost all Bayesian statistical procedures are, essentially, extensions of this way of thinking. Moreover, Bayes' theorem may be applied repeatedly, taking into account new statistical data and assuming now $P(H|E)$ (i.e., the *final* probability "of yesterday") as the *initial* probability "of today". And so on: there is no distinction at all between initial (or *prior*) and final (or *posterior*) probabilities, except that *they refer to a different state of information.*

In the so-called "ortodox" approach to statistics, there is no possibility to speak of "probable" conclusions concerning events such as H, just because this would involve probability as "degree of belief" and not as a frequency. So the attempt is made to introduce concepts such as that of *accepting* or *rejecting* a given hypothesis H, on the basis of the conditional probability $P(E|H)$ only: for example, if this probability (in the sense of observed frequency in the past) is "very large" (i.e. E is highly probable *given* H) and E in fact occurs, this may be considered as a good reason for accepting H. Notice that this argument corresponds to *ignore* not only *other* hypotheses but also the *initial probability* P(H); moreover, the acceptance of the hypothesis H is not based on any judgement of plausibility, but it is a sort of a mechanical act, based on the frequency of success of the method on which the acceptance criterion is based. Accepting (or rejecting) a hypothesis is a concept that after all is alien both to the *logic of certainty* (i.e., the usual, ordinary logic) and to probability theory. Nervertheless *an interpretation in Bayesian terms is possible* assuming *equal* initial probabilities of all possible hypotheses and also when an experiment can be performed a large number of times, so that the observed data do influence the evaluation of probability in a prevailing way as compared to initial probabilities: of course, in both circumstances the statement "*accepting a hypothesis*" may be … accepted only as a way of saying that the given hypothesis has a "very large" subjective probability.