Int. Stat. Rev., Vol. 42, No. 2, 1974 ,pp. 117–130/*Longman Group Ltd/Printed in Great Britain*

Bayesianism: Its Unifying Role for Both the Foundations and Applications of Statistics [1]

Bruno de Finetti
Rome, Italy

Summary

A fruitful discussion on the subject requires, first of all, to disentangle it from the difficulties arising whenever single aspects are considered separately, as unconnected technicalities concerning isolated problems. Induction is viewed usually in many very different, partial, unsatisfactory, isolated fragments of theories. There is the formalistic one of a kind of logicians; there are the two mathematical ones concerned respectively with inductive reasoning and inductive behaviour; and all became more and more complex owing to a growing inflation of formalism. Formalism is, in effect, the illusory remedy to the insufficiencies arising from isolation, and is in turn a factor of enhancing isolation.

The thesis of the present paper goes just in the opposite direction, trying to show, as clearly and simply as possible that, avoiding the misleading preconceptions of artificial unilateral constructions, the whole subject admits a unique and very natural interpretation and a simple universal answer. That is the Bayesian theory: but it is a pity it is called a "theory" and has a name, for the same reasons that led Cornfield to say (noting that Bayes' theorem is but an obvious result) that "it is overly solemn to call it a theorem at all".

Probabilities have a unique true meaning as degrees of belief (a subjective one, although one must take into reasonable account the objective data available: symmetries, frequencies, analogies, from all his experience), and must coherently agree together and coherently evolve by changing of the state of information. (That is, summarizing, the Bayes' theorem.)

All that is imposed, in a unique way, by concordant reasons pertaining to each of the aspects (and are, in their essence, an unique reason which presents itself under slightly modified appearance in the various occurrences).

These views, which in the present summary could only be sketched in an abstract and apodictical form, are carefully exposed in the present paper, comparing the effect of conformity to them, or of deviations, on all possible facets of the problems concerning induction, as for logical conclusions and for inductive reasoning and behaviour. Objections against subjectivity should be overcome; abstaining from subjective opinions yields in fact no improvement on objectivity; on the contrary, a subjectivistic integration to the barren bulk of the objective data appears to be necessary, so that, if it is missing, we are led not to any better situation but to a more erratic one.

1. A Fundamental Distinction

In order to avoid ambiguousness it is necessary to note, at once and firmly, an essential distinction between what may be respectively called *Bayesian standpoint* and *Bayesian techniques*.

Both apply to the problem of statistical (or probabilistic) inference, and are, of course, perfectly connected together as being the logical framework and the mathematical tool of the same theory: namely, of the theory concerning the way in which our opinions (or beliefs) must be modified (according to Bayes' theorem) when new information is attained. Nevertheless, in practice, the overlap of the fields of the published applications inspired to the Bayesian standpoint and of those making use of Bayesian techniques seems rather narrow.

In fact, most applications of Bayesian standpoint in everyday life, in scientific guessing, and often also in statistics, do not require any mathematical tool nor numerical evaluations of probabilities; a qualitative adjustment of beliefs to changes in the relevant information is all that may be meaningfully performed. And conversely, Bayesian techniques, more or less

[1] This was an Invited Paper at the ISI 39th Session and is reproduced here by kind permission of the author.

developed into imposing mathematical machinery, are often applied as such, using standardized "models" and standardized "prior distributions", instead of carefully keeping realistic adherence to the specific features of each particular case and to the true opinion of the person concerned (the statistician himself, or the decision maker, or somebody else).

Thus, the given distinction, fundamentel *per se*, is also necessary for a preliminary explanation of the thesis maintained in the present paper, and of the succession of the aspects that will be considered and discussed here in order to clarify the point of view defended.

To begin with, let me express it very roughly as follows.

Bayesian standpoint is noways one among many possible theories, but is an almost self-evident truth, simply and uniquivocally relying on the indisputable coherence rules for probabilities. It should be always applied in the most natural and naïve form, paying attention – whenever the recourse to a more sophisticated machinery seems unavoidable – that its introduction should not induce to lose sight of the true situation and opinion.

At the contrary, Bayesian techniques, if considered as merely formal devices, are no more trustworthy than any other tool (or *ad hoc* method, or "Adhockery" to use the word introduced by Good) of the plentyful arsenal of "objectivist Statistics". It is true that Bayesian techniques give rise to all (and only all) the *admissible* decision rules (according to Abraham Wald), but each one is valid with reference to a particular initial opinion; therefore, any conclusion is arbitrary if the choice (purposedly or unadvertedly) responds to arbitrary formal criteria (inspired, e.g., to simplicity, or to mathematical convenience) rather than to personal advice. And, moreover, two or more admissible decisions may constitute, together, an unadmissible compound decision when based on incompatible initial opinions instead of on the same one.

2. The Essence of the Bayesian Standpoint

The essence of the Bayesian standpoint can be better and more clearly grasped when discussing it with reference to very simple but practically meaningful applications.

It may surely be convenient, for instance, to consider only the case of a finite (or even small) number of possible "*hypotheses*" and of available "*decisions*". This is, no doubt, a severe simplifying assumption, but a fundamentally innocuous one, since it allows to keep in due account, at least approximately, all the relevant features of the particular problem one is facing.

At the contrary, any reference to general probability spaces,[1] parameter spaces and decision spaces would be inappropriate and often misleading for the present purpose, because the attention would be attracted to the overwhelming framework of the analytical machinery, where the conceptual aspects are incorporated in a way that often conceals the true meaning of the assumptions. Moreover, the choice of such framework appears, in some kinds of exposition, a largely arbitrary adoption of one among a collection of standard cut-and-dried models rather a careful effort to truly represent a reasonable state of doubt (of partial knowledge and of educated guessing in face of ignorance) about our present problem in the present state of our information.

I found very illuminating, as for the essence and irrefutability of the Bayesian standpoint, the explanations and exemplifications in the paper on "Bayes Theorem" (Cornfield, 1967) and in the booklet *Making Decisions* (Lindley, 1971). They are nowise carrying new proofs or new arguments on the subject, which seems in itself completely clear and simple, but they appear particularly fit to fill the psychological (or, in a sense, didactical) need of dissolving the intricacy of preconceptions which hinders an open-minded debate. To make frequent use of quotations from such works should be, I hope, justified, inasmuch as they express what I

[1] It should also be noted that, when a probability distribution in such spaces is considered, it is usually admitted it obeys σ-*additivity*. That is – in my opinion – an unjustified and even untenable assumption. However, this question is not of basic importance here, and will be ignored.

would say surely better (at least as for English, and probably in absolute); my additions or comments will at any rate specify, when necessary, some nuances.

The contrast mentioned between simplified but honest methods of comparing decisions, and sophisticated but perplexing ones, is happily illuminated by Lindley:

In our everyday decision-making we[1] have developed, because there were no basic rules to guide us, some bad habits. One of them is the tendency to shy away from the simple and take refuge in the complex, where *it is not so easy* to have *one's incompetence exposed.* As has been said: "Practical decision-makers instinctively want to avoid the rather awful clarity that surrounds a really simple decision". The reply to the accusation of guessing at probabilities and utilities is simply that if you can't do simple problems, how can you do complicated ones? (pp. 64–65).

The material in this book provides a tool *to aid* the decision-maker: it does not try *to replace* him; . . . we do not offer a machine whose handle only needs to be turned to demonstrate the proper course of action: we merely provide some *guide lines* for *sensible* decision making, guide lines which enable a complicated decision process to be broken down into smaller, and therefore simpler, parts whose separate analyses can be combined to provide a solution to the whole (p. 2). The framework does not require anything sophisticated in the way of mathematics, though it does require a little of logical abstraction (p. 180).

Essentially, this book is about *coherent decision-making.* – We shall not study how decisions are made today. – Many studies of these types . . . appear to show that man does not make decisions in accord with the recipes developed here: in other words, he is incoherent. Such empirical results reinforce our belief that *the statistician's contribution is significant*: he appears to have something new to say; he is not confirming man in his present ways. Our method is not empirical; we sit back and think about the decision process, and *show* that it *must* have certain features (pp. 3–4).

3. Admissible Decisions are Bayes' Decisions

How to *show* that the decision process *must* have certain features, and precisely the Bayesian ones? That has been repeated and explained by all Bayesian authors; the essential point consists only in finding the key to convey the true meaning of a thesis in such a manner that its interpretation should not be distorted or hastily refuted owing to counteracting preconceptions. Hopefully, this aim should be fulfilled by the Ariadne thread formed joining together some quotations from Cornfield's paper; his elementary but deep and cogent exposition, even if summarized by extracting the leading sentences, should indeed be sufficient to present and prove the necessary character of the Bayesian foundations for inductive reasoning and (consequently) for inductive behaviour.

A set of observations – he remarks at the beginning (p. 34) – may be logically consistent with several hypotheses [or "states of nature"], even though some of the hypotheses are inherently less plausible than others and even if the observations are more reasonably accounted for by some hypotheses than by others. We all characteristically draw conclusions in such situations and these conclusions guide further thinking and research and influence our behaviour. The conclusions drawn are uncertain, however, so that it is reasonable to seek some quantitatively consistent way of characterizing this uncertainty. Bayes' theorem is important because it provides an explication for this process of consistent choice between hypotheses on the basis of observations and for quantitative characterization of their respective uncertainties (p. 34).

Bayes' concepts have been enormously broadened and deepened since 1763. It is now possible that the theorem *is not just another possible explanation of inference and decision* but that, if a simple unified explication is possible at all, then there is a precisely defined sense in which it must be consistent with Bayes' theorem (p. 34).

Actually, Bayes' result follows so directly from the formal definition of probability and related concepts that it is perhaps overly solemn to call it a theorem at all. – We emphasize that although there are differences of opinion on the extent to which Bayes' theorem can be applied, the theorem itself is universally accepted as mathematically correct, so that none of the following [points] should be subject to controversy (p. 36), . . . [except for] the assignment of . . . the unconditional or prior probabilities of the various states [which] presents more difficulties in principle [owing to the controversy about the purposiveness of a] distinction between frequency and non-frequency interpretations in probability (p. 40).

Nevertheless, the work on Decision Theory, as developed, chiefly by Abraham Wald (1939), according to an *objectivistic* point of view, led to distinguish *admissible* and inadmissible decision rules, and to state that the admissible ones are all and only all those that maximize

[1] Here "we" refers not to Lindley's own attitude, but to the one of the majority.

expected utility[1] *according to any possible set of initial probabilities*. Admissibility is indeed defined in a very obvious sense: a decision rule has to be discarded as inadmissible if it is *dominated* by some other decision rule, that is, if another decision rule yields an outcome which is better whatever happens (or, at least, always better or sometimes equivalent).

But there is not a *best* rule, a rule dominating every other, which would be then the only admissible rule. So that we are only obliged to

> confine our choice to admissible decision rules. Anyone using an unadmissible rule will incur a greater average loss for at least one state of nature than would be incurred by using any of the admissible rules which dominate it and will not incur a smaller loss for any state of nature. There seems to be no possible construction of the word "best" that would include an inadmissible rule (p. 43).

That can be expressed, as vividly as did Lindley (pp. 21 and 58), saying that "the incoherent person (as a probability assessor or as a decision maker) is a *perpetual money-making machine*" for advised opponents.

The conclusion is straightforward.

> The Bayesian can thus justify [toward an Objectivist] the assignment of prior probabilities and the use of the rule by the fact that it is *sufficient* for admissibility. But someone who is unwilling to assign prior probabilities might inquire about the *necessity* of the Bayes' decision rule. Are there admissible rules which are not Bayes' decision rules for any possible set of prior probabilities? The answer is a clearcut no, [and] we see that the use of a Bayes' decision rule relative to some set of prior probabilities is *necessary* for admissibility.
>
> This result places anyone who accepts admissibility but denies the existence of prior probabilities for unknown states of nature in an awkward, if not untenable, position. If he produces a decision rule which is admissible, it is a Bayes' decision rule relative to a particular set of prior probabilities. His preference for this rule is thus formally identical with the assignment of these prior probabilities (p. 43).
>
> Although these results do not indicate how prior probabilities should be assigned they do indicate that reasonable behaviour is equivalent to their assignment and conversely. That makes the revival of interest in Bayes' understandable, and suggest that it may be more than a passing fancy (p. 44).

4. The Opposition to Bayesian Standpoint

Nevertheless, some people refuse to assign initial probabilities, that is to assign probabilities to "hypotheses"[2] (or, equivalently, to the possible values of a parameter θ that could be introduced to distinguish them) because θ is an "unknown constant" (not a "random variable" understood as something assuming different values at random in a series of "repeated trials"). So, probability does not admit here any frequency interpretation; it is overtly a subjective probability, and that is anathema for strictly objectivistic statisticians.

They seem so superstitiously terrified by such anathema that they usually prefer inadmissible decisions based on a variety of *ad hoc* rules, either rough or sophisticated, rather than an uniformly better (and admissible) one obtainable by Bayes' method. Cornfield remarks somewhat humorously:

> The idea that prior probabilities need not be frequencies is considered by some frequentists to be related to "those absurd conceptions of non-empirical, *a priori*[3] known probabilities that cannot be tested by any experiments, etc. This cannot be strongly enough refuted" (von Mises, 1942). Since θ is usually (if not always) an *unknown constant* the frequentist in practice usually rejects Bayes and on occasion finds himself using [inadmissible] decision and inference procedures. To the non-frequentist the use of such inadmissible procedures seems like *an extravagant price to be paid for the support of a philosophicai position, and an empirically unverifiable one at that* (p. 46).

At this point we cannot defer further the discussion concerned with the validity and purposiveness of the distinction between frequency and non-frequency interpretations of probability, which "constitutes the great point of controversy in the application of Bayes' theorem"

---

[1] We do not discuss the notion of *utility* here. It is the now usual one, as, e.g. in the Appendix of von Neumann and Morgenstern (1944). As for our arguments, it would be indifferent if one had monetary value as his own utility.

[2] See, however, sections 8 and 9.

[3] To refuse *a priori* probabilities is right. But our probabilities (initial or "prior", final or "posterior") are always *subjective* (not *a priori*).

(p. 40), either in general or in the specific case of this assignment of initial probabilities. There are, however, several more questions to be discussed later, in order to improve the presentation of the Bayesian framework too (maybe, incidentally, with the effect to make some of its aspects somewhat more palatable to Objectivists).

Let us begin with a few more quotations from Cornfield:

> The inability to assign in a unique way prior probabilities either from experience or from principles like that of ignorance or invariance has led to a favourable re-examination by some of the nineteenth-century doctrine as expressed by De Morgan (1847) that probability is a degree of belief. De Morgan held that *a probability is not an objective characteristic of the external world, but a subjective attitude towards it, which can and does vary from individual to individual.* – If one accepts such a view of probability one must reject the idea that an outcome must lead to, in Fisher's words, "a rigorous and unequivocal", i.e. a unique conclusion.
>
> I must confess – says Cornfield – that although I once entertained objections somewhat like this *I now regard the subjective view as inescapable.*
>
> To statisticians who have taken a somewhat authoritarian viewpoint about their ability to draw unique conclusions from small bodies of data and to design experiments best able to produce such data this conclusion is unacceptable. To those who have always doubted the possibility or desirability of eliminating personal judgment in the design or interpretation of experiments the existence of an orderly mathematical way of combining prior opinion and evidence should prove welcome (p. 47).

All that is completely in agreement with my own point of view, with the sole proviso that what has been said here for *prior* probabilities applies to *all* probabilities. There is no distinction at all between initial (or prior) and final probabilities except that they refer to a different instant (and, then, to a different state of information, etc.); or, in other cases, they may differ as representing the opinion of one person or of another one, and so on.

The quotation shows also that my conclusions (reached in the 1928–1930, almost as soon as I met the notion of probability and discarded it because of inconsistency of the usual "definitions" based on symmetries or frequencies) were not new, but (as I did learn later) coincide with the ideas of De Morgan (1847) and of Ramsay (1926).

Such ideas are however distressing for some people, who consider objectivity, in the strictest sense, as a necessary attribute of probability and of science. But the regret for losing the faith in the perfect objectivity of probability, and hence of science, is unjustified. Nothing is lost but what was a mere illusion. And such illusory objectivity is now replaced by the effectual objectivity, that is the true degree of objectivity attainable by human science through human senses and mind. Cornfield says:

> The objectivity of science finds its mathematical expression in the fact that individuals starting with quite different prior probabilities will nevertheless compute essentially the same posterior probabilities when faced with a sufficient large body of data. [That corresponds also to a methodological conclusion quoted from Mosteller and Wallace (1963): "Prior distributions are not of major importance. While choice of underlying constants (choice of prior distributions) matters, it doesn't matter very much, once one is in the neighbourhood of a distribution suggested by a fair body of data. We conclude from this that the emphasis on the difficulty, even impossibility of choosing prior distributions as a criticism of the use of Bayes' theorem is not well placed."]
>
> Whether one eventually accepts or rejects this conclusion, it is clear that *it is not possible to think about learning from experience and acting on it without coming to terms with Bayes' theorem* (p. 47).

5. Frequency and Non-frequency Interpretation

To introduce the notion of probability, reference is usually made to the notion of frequency. Why? Cornfield says "this is more or less a historical accident" (p. 37); but, again, why? Why should such an awkward accident so pertinaciously endure?

My answer is a very simple and natural one, in my opinion, although most people would consider it paradoxical: *The persistency of the frequency interpretation of probability hangs on its being the worst possible one.*

It is the worst because there are so many distinct connections between two notions so reciprocally alien in their essence as probability and frequency are, that, since a confusion has occurred, any effort to clarify the situation escaping the unnatural identification risks being ineffective, like in the farces where identical twins are continuously interchanged giving rise to funny and absurd misunderstandings.

Such confusion is so much more difficult to overcome because of the terminological trick, unfortunately so widespread, of calling "events" not the single events but some vaguely defined "species" of which the single events are "trials". This way, the rich and nuanced variety of analogies, similarities and dissimilarities, possible correlations and so on between the single events, are carelessly ruled out. Each author feels justified in considering the events of a class whatsoever as equally probable, and (if convenient) stochastically independent, as soon as he introduces a name for a "species" to which they are said to "belong" as "trials"; and, although his definitions of notions like independence are only valid and meaningful for the entities *he* calls "events", and meaningless for *our* events that *he* calls "trials".

In contrast with such objectivistic jargon, and according to the subjectivistic views, every probability one may be interested to appraise (as well as every probability altogether) refers to a *single, well specified, event:* the shipwreck of the vessel we are considering to insure for its next voyage; the diagnosis and prognosis for this particular patient under such treatment or another; the success of a given candidate in an election – of a particular student in passing a specified examination – of a given football team in the next match; and so on. This remark is noways intended to deny the usefulness of considering and confronting any single event of interest in connection and comparison with others, more or less similar under any of the possibly relevant aspects and usually called with a common name (like "a head by coin tossing"); but every relevant circumstance must be realistically pondered case by case.

Frequencies may enter into this framework only incidentally, although in several different ways. Observed frequencies of past outcomes in events more or less similar to the one (or ones) of present interest must be taken into account because they concur to the present assessment through Bayes' theorem, as every other piece of information. (This point becomes significantly illuminated when related to the notion of exchangeability: see section 9.)

Future frequencies (or, indifferently, past ones, but not observed or not known to us) may also be suitable ingredients for the analysis of our problem, and that at least in two ways. On the one hand, it may be easier (either by applying a mathematical setup, or by educated guessing) to estimate a related frequency rather than directly the required probability. On the other hand, it may happen that a frequency is the variable of interest needed for the question, so that it is chiefly worth while to know its probability distribution (or – whenever that seems sufficient – its expectation, or its median, or any other similar characteristic).

Of course, in both cases the specific circumstances of the question considered must be taken in due account. For instance, it is mistaken (although rather usual) to think that it suffices – in order that the events of a given class (or "trials of a given event") should be considered "stochastically independent" – that no direct influence at the outcome of one of them acts on the others. There are, in fact, also indirect sources of interdependence, owing to the possible existence of causes influencing all (or many of) the events considered, or owing to changes in the state of information (see, e.g. de Finetti, 1970*a*, pp. 178–184).

6. The Assessment of (Subjective) Probabilities

It does not belong to our present task to discuss the validity of the notion of probability (from the subjectivistic point of view), and the methods by which it may be assessed. It is unavoidable however to touch, at least shortly, this point, in order to prevent misinterpretations.

Probability as *degree of belief* is surely known by anyone: it is that feeling which makes him more or less confident or dubious or sceptical about the truth of an assertion, the success of an enterprise, the occurrence of a specific event whatsoever, and that guides him, consciously or not, in all his actions and decisions.

It suffices, ordinarily, to know such degree of belief qualitatively, but its measurement by a quantitative scale is widely understood and applied as basis for economic operations under uncertainty: bettings, insurances, risky investments, expenses for protection, and so on. Coherence in such operations (admissibility; avoidance of Dutch Book) requires coherence in probability assessment (what is tantamount to accept the principles of probability theory).

In particular, if we find a partition in (e.g.) 100 cases to which we assign equal probabilities (namely, of course, 1 per cent) we get a scale apt to let us assess probability to every event by direct comparison: it is, e.g. between 16 per cent and 17 per cent if it is more probable than an union of 16 (but not of 17) of the 100 cases. (An equivalent comparison could be presented with a scheme of independent repeated trials with expected frequency between 16–17 per cent, but that would involve several ill-definable notions before probability itself.)

Direct recourse to betting may be improved by constructing devices that punish deviations; such methods are noticeable also for applications to psychological experiments on the subject (see L. J. Savage, 1971, and B. de Finetti, 1970*b*).

But often the real significance of probability, even for people prevented because of preconceptions from expressing it by a number, is easily disclosed making him express indirectly his own assessment through a practical decision. Here is an example from Lindley:

> A chemical engineer realized that there was a chance of the process for which he was responsible failing but was reluctant to assess it numerically. However he knew the monetary consequences of failure and so I asked him: suppose I was able to offer you a device which would make the process certain, how much would you pay me for it, a thousand dollars, ten thousand? He laughed at the latter figure as being ridiculously high but contemplated the former more seriously. After some bargaining we settled for 750; a figure which can be converted into a probability (given the loss) (pp. 25–26).

Another important way for improving an assessment of probabilities consists in extending it to other events according to the laws of probability, and to check whether all the probabilities obtained agree with our true opinion.

> If this deduction leads to values that seem unacceptable then we must revise some at least of the original values to reach a set which both agree with our ideas and obey the laws. – The laws provide means whereby many probabilities can be calculated in terms of some basic values, [but] *no probabilities are any more basic than any others* (p. 38). Probabilities should be not judged in isolation: they should be compared, one with another (p. 44).

Similar considerations may also bring different persons to reciprocally approach their opinions; besides, also a rough qualitative appreciation of probabilities *is* sufficient and worthy for practical purposes.

7. Importance of (at least) Qualitative Bayes' Inference

Inference deductions or conclusions do not often need quantitative precision, but surely a qualitatively correct adherence to Bayesian requirements. A few disparate examples may suffice to illustrate this point.

Plausible reasoning in pure mathematics consists in assigning (more or less qualitatively) a probability to the truth of a supposed theorem, to the success of endeavours to solve a given problem, and so on. An effort directly intended to quantify such belief would surely be idle in itself, but this belief underlies to the mind and behaviour of anybody engaged in any research or problem, inspiring more or less confidence, and hence stimulus, to progress.

Polyà (1953) discusses *Patterns of Plausible Inference* (this is the title of Vol. II), from which we quote some remarks by way of example:

> The direction [of a change in opinion after a consequence of an hypothetical theorem has proven to be true] is expressed and is implied by the premises, the strength is not. – The direction is impersonal, the strength may be personal (p. 114). – The verification of a consequence renders a conjecture more credible: The increase of our confidence in a conjecture due to the verification of one of its consequences varies inversely as the credibility of the consequence before such verification (pp. 120, 121). – Our confidence in a conjecture can only diminish when a possible ground for the conjecture has been exploded: The more confidence we placed in a possible ground for our conjecture, the greater will be the loss of faith in our conjecture when that possible ground is refuted (p. 123). – [Conversely in the case] when an incompatible rival conjecture has been exploded (p. 124).

After an attempt at numerical computations for a particular problem (exposed to criticisms, as himself admits):

. . . we may find it safer to return to the [former] standpoint: . . . represent to yourself qualitatively how a change in this or that component of the situation would influence your confidence, but do not commit yourself to any quantitative estimate (p. 132).

Such conclusion is somewhat too pessimistic: while too precise estimates are obviously silly, a reasonable guess on the size (e.g. 20–30 per cent, 0·5–1 per cent, 10^{-7}–10^{-6}) is valuable.

A rather analogous kind of question is the one concerning other scientific conjectures. Let us only mention a paper by Good (1969) about Bode's law (about the distances of the planets from the sun): is this regularity due to chance, or does it depend on some scientific ground? The research is interesting in itself and as a novel application of Bayes' methods, but still more for a bitter ensuing discussion. It is incredibly strange how the honest way of trying to express numerically his own degree of confidence has been attacked, whilst, generally, every superficial and unexplained preference of an authoritative scientist suffices to guarantee acquiescent consent to one theory and a blind disdain for the rival ones. That seems terrific as a token not only of scientific irresponsibility but also of a widespread bluntness exposing mankind to any threatened evil (see also, de Finetti, 1971).

Reverting to more properly statistical research, a few examples should suffice to defend and to illustrate the same attitude. The best decision about a journey (in Lindley, 1971, pp. 32ff.) depending on a pass being perhaps blocked by snow, possible accidents, etc., implies obviously a subjective appreciation of several probabilities, and their adjustment to new information. As, e.g. a friend tells the interested people: "Yes, it has been like that in the past, but the local authority has . . . (done so and so) to keep the pass free of snow", *this additional fact* made him naturally revise his probability downwards. And Lindley adds:

There is nothing in our argument that makes agreement inevitable, but in practice it will often happen that agreement can be reached, given enough evidence. This is one reason why information is a good thing (p. 32).

In the same spirit a Bayesian (or, at any rate, myself) must feel sympathetic with similar attitudes even if not technically Bayesian: e.g. with the views of J. W. Tukey (1962) about Data Analysis. Among many consonant possible quotations, here is one:

The most important maxim for data analysis to heed, and one which many statisticians seem to have shunned, is this: "Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise" (pp. 13–14).

Also some critical remarks about Bayesianism by Egon Pearson (1962) appear to me wise recommendations for a sensible use of Bayesian standpoint rather than rejection of its use outright. As an example (only one, to save space) of good-sense recommendation from users of statistics in cognate field, here are some fragments from D. B. Suits (1967) on Econometric Forecasting:

. . . people who actually forecast are rather relaxed about standard errors and about such questions as least squares bias in systems of equations. They are only too well aware that the big problems lie somewhere else entirely. – Analysis of forecasting failures is the greatest single source of new knowledge and insight (pp. 235, 246).

8. How to get rid from the Framework of "Hypotheses"

To get rid from such framework is the aim explicitly expressed in H. V. Roberts (1965) advocating that attention should be focused on the "*predictive distribution*" concerning directly the quantities of interest rather than the ones concerned with the "parameters" that are but auxiliary ingredients.

A simple example – presented in two versions – should suffice to make the issue clear: it is the well-known Bayes-Laplace scheme of events "independent and with the same probability θ, the unknown constant probability θ being uniformly distributed on 0-1". To get an approximate but more realistic idea, one may perhaps think of independent drawings from an urn "chosen at random from among 1,000,001 urns, each one containing one million balls, of which the white ones are respectively $h = 0, 1, 2, \ldots, 1{,}000{,}000$; the proportion of white balls, θ, is h millionths, with probability 1/1,000,001 each.

After n drawings, with m occurrences of white, it may be of interest to know that now θ, the unknown proportion of white balls, has the Beta distribution with density $K\theta^m(1+\theta)^{n-m}$, but it is probably more interesting, for practical purposes, to know that the probability of white in any future drawing (as estimated now) is $(m+1)/(n+2)$ (Laplace's "rule of succession").

But consider now the Polya's urn scheme (of "contagious" probabilities): there are initially in the urn just 2 balls, 1 white and 1 black, and, after each drawing, the drawn ball *plus another one of the same colour* is put into the urn. After n drawings with m occurrences of white we have then in the urn $n+2$ balls, $m+1$ of which white. The two cases considered are but different versions of the same abstract process.

From a realistic point of view there is, however, an essential difference. In the first scheme, θ is a factual, although unknown or "*hidden*" quantity; one could check its value if only it were not forbidden to inspect the content of the urn; in the second, θ is a merely fictitious, or "*mythical*", pseudoentity, allowing to perform *some* reasoning *as if*[1] it "existed", but leading to absurdities were it to be thought as really "existing".

Some old-time demographers, in the age when probabilities were ordinarily conceived only with reference to urn-schemes, used to explain the mortality table with the image of the Parcae drawing each year a ball for each of us, to decide about life or death according to its colour – white or black – and using an urn where the fraction of black balls was increasing with the age.

But to imagine an urn with unknown but constant composition explaining at any drawing from the Polya urn its outcome as resulting from the "hidden urn" would be even more difficult: much more artificial and preposterous than the plainly mythological picture of the Parcae. The "hidden urn" should in effect have, as its unknown but *predetermined* composition, the one which corresponds to the limit to which the composition of the Polya-urn should approach through endless additions of new balls to the few put into it till now. (And it is almost sure that not even the Vestals would assure the continuation of such experiment for the eternity, what would imply, incidentally, to get sometime more balls than atoms in the world; and, on the other side, there is no reason to expect such limit *exist*, since "stochastic" [even if strong] convergence does not guarantee any conclusion on this point.)

In such situation, it is obviously only the *predictive* aspect (concerning the *future outcomes* – not the parameter!) that matters.

9. Exchangeability

A particular case (but a particularly simple and important one) of Roberts' distinction is the one where (as in the example just discussed) we have – in the terminology of Objectivists – "independent" events E_i (or random quantities X_i) with the constant but unknown probability θ (or probability distribution, say F_θ). Under such assumption, the E_i (or X_i) are *exchangeable*: i.e. any event depending on (distinct) events $E_{i_1}, E_{i_2}, \ldots, E_{i_n}$ (or random quantities $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$; e.g. $E =$ among the E_{i_h} just m occur; or m of the X_{i_h} are $\leqq k$; ...) has the same probability no matter what E's or X's are chosen. The probability distribution is, in other words, symmetric.

But the converse also holds: *exchangeability* implies the possibility (usually but as a "mythical" interpretation) of the said formulation in the terminology of Objectivists. That is what some Colleague (I don't know who and when; I noticed but late that this denomination

[1] Remember the "*als ob*" of Veihinger's philosophy.

was rather common) called the "de Finetti's representation theorem". My aim was precisely the same as Roberts', if only restricted to the case of the most usual and important example of inductive reasoning (and, then, of inductive behaviour). The aim was to present induction as a very natural way of reasoning on probabilities of observable facts avoiding metaphysical, pseudoentities and obscurities.

10. Inductive Reasoning and Inductive Behaviour

It is advisable, having mentioned Inductive Reasoning *and* Inductive Behaviour, to recall briefly the often believed different idea of an opposition of Inductive Reasoning *versus* Inductive Behaviour. On this point, even starting from contrasting conceptions, two outstanding leaders of modern statistics as Ronald A. Fisher and Jerzy Neyman shared the same attitude, which is still widely supported.

The conclusion arrived at in our present discourse (as, of course, in earlier work of all Bayesian authors) solves the question by identifying the two processes, in the sense that an admissible inductive behaviour is the one which maximizes expected utility of any decision problem according to an admissible utility of any decision problem according to an admissible inductive reasoning; which in turn requires starting from probabilities expressing our initial opinion and coherently to revise them taking into due account every subsequent observation or information according to Bayes' theorem.

This way, it should appear impossible for everybody to admit any different solution. In fact, statistical decisions now appear reduced to a simple and obvious comparison between different itemized accounts (or "bills"): any uncertain gain or loss (of utility; for moderate amounts also the monetary value may suffice) has to be multiplied by its price, that is by its probability, like in any insurance policy. There is nothing else to do, nor to excogitate, in both cases. Even if somebody finds the true rule for accounting and for decisions despicably *dull* (as it did happen) it seems unwise *to pay the extravagant price* (in Cornfield's words; see quotation from his p. 46 in our section 4) even for the pleasure of using an opinionatedly nobler or fascinating or fashionable mathematical machinery. (As a matter of fact, nobody did suggest, *for bills*, any "better" [nobler, or more fashionable] rule than the old and trite one of summing up the products of quantities by prices. Why? Are businessmen – with their apathy toward seeking such kind of novelties – more backward or more skilful?)

It would seem unnecessary to specify in detail why and how any deviation from the Bayesian approach leads to mistakes; once we learn that *Bayesian* and *admissible* are synonyms, we could suppress every attribute tacitly assuming we are not interested in inadmissible, non-Bayesian rules. However, that cannot be done prematurely, and, moreover, there are always instructive lessons to learn by scrutinizing the specific effects of any deviation. It is that which we are willing to do, at least shortly, before concluding this review paper.

11. Differences in applying Bayes' Methods

Let us begin with the applications of Bayes' methods: they are themselves open to distortions, chiefly owing to misunderstanding about the choice of the initial probabilities.

The choice must express our true opinion, or somebody else's opinion, or a sample of different opinions, specifying the case and the reason (this "somebody" is the decision-maker, his adviser, or so; a sample is used in order to have and to give of how conclusions are varying with the premises, and so on).

Is it admissible to choose the initial probabilities, or to modify their choice, after comparison and scrutiny of the consequences? Any choice or change with the aim to get a decision preferred for other reasons or interests would be obviously incorrect and usually dishonest. A different situation may however arise, and lead to a different appreciation, if some (preferably simple) inferences are explored in order to check what the initial opinion really means, which at first sight appeared best suited to express our true opinion. Let us remember (quotation Lindley's p. 38, end section 6) that "no probabilities are any more basic than any other"; so a guess based on some consequences may appear more reliable than the one concerning the initial conditions directly. Let us think to the example (often mentioned by Good and Savage): the prior probability that a person has some extraordinary quality, as he asserted, may be better grasped by asking ourselves after how many consecutive successes the posterior probability should attain one-half.

To choose simple distributions for initial probabilities (uniform among a finite number of hypotheses, uniform or Normal or other usual ones for continuous parameters) is rather innocuous if that is a simplifying assumption qualitatively adherent to our true opinions; otherwise not. (Beware that "uniform" – or "normal", and so on – does not *per se* convey any significant – or, in a sense, "natural" – meaning: every distribution is uniform, or normal, etc., if only we assume another parameter, $\theta' = g(\theta)$, just as much arbitrarily chosen at beginning. For example, with $\theta' = F(\theta)$, F the distribution function for θ, θ' has uniform distribution in [0,1]).

At any rate, the initial opinion concerning a set of hypotheses *must* be the same if used repeatedly for different inferences and/or decisions: in particular, it must not be changed (explicitly or inadvertedly) making it dependant on the nature or value of the losses (or gains) associated with different outcomes, and so on. This fact ought to be emphasized, both because such mistake sometimes also occurs in works by Jeffreys – one among the most prominent leaders of Bayesian standpoint – and because that has been objected to as an inherent mistake of Bayesians (not an occasional one of some of them). Such an idea seems necessarily to underly the distinction insisted between "scientific inference" and "practical decisions" by several authors (like Neyman and Fisher; probably that is also supposed by Pearson, 1962, pp. 397, 401).

Adopting different initial opinions for different problems depending on the same set of hypotheses, the ensuing conclusions and decisions constitute together an inconsistent conclusion and an inadmissible decision. In particular, the minimax method is inadmissible unless it leads by chance, for a single decision problem, to the initial distribution that corresponds to our opinion. That cannot, however, be true for more distinct decision problems depending on the same hypotheses, so that joint conclusions "explode".

As for the minimax approach some authors seem to consider it a suitable method when risk aversion is strong; but that would only imply a stronger convexity of utility, with probabilities unchanged. To apply indifferently a "more severe" evaluation on the side of probability or of utility, disregarding what one is really to be re-estimated, is an inadmissible slip, and an inexplicable one unless by the formalistic attitude often unfortunately induced by the habit of applying rough *ad hoc* devices. (With reference to the business example, that is tantamount as suggesting that, instead of increasing prices, one could perhaps reduce the length of the *metre*, disregarding the disarray produced on the price system because of the different units of reference [proportional to metres, to square metres, to cubic metres, or invariant toward lengths] as for wires, areas, liquids, time.)

12. Miscellaneous Remarks about non-Bayesian Differences

(a) *Refusing initial probabilities* implies that the hypotheses to which reference is made are said to be the only "possible" ones. It is likely that authors include in their list only the sufficiently "probable", discarding the less probable; but that is much more arbitrary and dangerous than to give them little probability. (Pearson [1962, p. 396] says in fact that, e.g. "the choice of the most likely class of admissible hypotheses" is done through an "intuitive process of personal judgement", that – it seems – differs not in being less subjective but only less rational.)

At any rate, using *ad hoc* (non-Bayesian) methods, it is almost sure one is lead to an inadmissible decision; if it is, by chance, admissible, one cannot judge whether the subjacent probabilities agree with his own subjective opinion (or at least with a neighbouring one). Sometimes methods imply inadvertedly (or at any rate tacitly) a uniform distribution (which may be far from any reasonable one).

Refusing initial probabilities is then not a way allowing any "more objective" inference or decision, based on "objective" data only; it is either untenable whatever the prior opinion might be, or it is equivalent to the blind adoption of a prior opinion that has only the dubious merit of being that one of a nameless and perhaps nonexistant Mr Somebody, preferred simply in order to shun the responsibility of a considered choice. Such choice need not indeed to be uniquivocal: one may reasonably discuss the consequences of several more or less reasonable opinions, and the pros and cons as for putting more reliability in this or this other opinion. To be completely neutral in judging choices, as by qualifying every choice as "arbitrary", seems an extremely strange and untenable position.

(b) "*Empirical Bayes*" is a seemingly intermediate method where the "possible parameter distributions" are known, but not the probability distribution over their set (or space). Robbins (1964) admits himself that such compromise is nonsensical both for Bayesians and non-Bayesians (and I was unable to grasp the point). I noted, however, from some passages that a kind of *predictive* process seems allowed (distribution of X_{n+1} given $X_1, \ldots, X_1$); if the order of the n observed X_h is indifferent (a fact which might look rather plausible, at first sight), the method would coincide in disguised form with the Bayesian one under exchangeability. This is, at any rate, only a tentatively advanced conjecture.

(c) *Some predictive rules* (like the Bayes-Laplace one, of $[m+1]/[n+2]$) are *accepted or proposed by logicians* as *ad hoc* methods of "inductive logic". It is hard to comprehend the rationale of such attitude, favouring uncritical excogitation (or acceptation as *deus-ex-machina*) of simple or complex, admissible or inadmissible, arithmetical expressions as "rules of inference" in the void, despising the natural and meaningful results concerning probability and possessing universal validity. Induction is but a particular and exemplar case, depending on nothing else as the application of Bayes' theorem. I think that this view agres also with the old but always illuminating ideas of Hume, although their lack of a quantitative formulation allows controversial theses to survive.

(d) *Accept or reject* is the unhappy formulation which I consider as the principal cause of the fogginess widespread all over the field of statistical inference and general reasoning. It is the same fogginess that would obscure a discourse if it were forbidden to mention the degrees of temperature (considered "unscientific") and obliged to "state" or "decide" whether the water in this glass is at the absolute zero or infinitely hot (either $-273{\cdot}1°$ or more than $100°$).

There is no reason (nor place) to repeat or summarize here what I discussed thoroughly many times, particularly in a recent book in English (1972). Let me only mention again sentences synthesizing the objections against the "accept or reject" dilemna: "We can invent something else to *say*, but nothing else to *think*" (Pratt, 1961), and "We cannot get a probabilistic omelette without using probabilistic eggs" (L. J. Savage).

13. On "More or Less Subjective" Probabilities

Wherever one could think as possible to place a boundary for the field where "objective probabilities" exist, their admitted applicability would be very limited in extent and usefulness. If the probabilities needed outside such a privileged field are consistently considered as "arbitrary constants", one could only conclude, with Boole (as quoted in Cornfield, p. 41), "that definite solution is impossible and to mark the point where inquiry ought to stop". I see no place for any intermediate position between that one and the subjectivistic interpretation.

Nuances are admissible only as for psychological preference to abstain from problems on "more subjective" probability assessment. Would such a distinction be admissible?

It has been suggested, and debated, whether such a distinction between "*more or less objective*" situations is possible (Blyth, 1972; Dempster, 1972). I agree with a change: I would say "*more or less subjective*"; this is not an idle subtlety: one may ask if a body is *more or less frangible*, not if it is *more or less infrangible*. (So, in my view, *objective* probabilities are only the values 0 and 1 for absolutely certain and absolutely impossible events or assertions.)

Quoting from Dempster:

For example, Blyth tacitly assumes that the exchangeability of his own coin spins is "objective" while the exchangeability of a mixed sequence including his own and somebody else's (all following the same rules) is "subjective". In fact, a careful observer would question both and would come away with a keen awareness of the subjective element in the decision to accept either kind or exchangeability.

All statisticians agree with the maxim "Let us look at the evidence". The evidence is: through what mode of thinking shall we look at the evidence?

That's right. Of course, every opinion is subjective and hence also every property admitted for it – as, in the example, exchangeability – is subjective. The (certain) fact that the coin is always the same gives to that opinion a lesser degree of subjectivity than when the coins are changed from some trials to others. In this case we might reasonably assume "partial exchangeability" (if we know what coin is used in any trial. If it were chosen at random at every trial and no information about that is available not even afterwards, any distinction fails.). About "partial exchangeability" see de Finetti, 1972, Sect. 9.6.2. and examples in Ch. 10 by Bruno, 1964.

That exemplifies a fairly nuanced attitude that may help subjectivity to be considered, as it must, not as something deliberately opposite to objectivity, but as the strictest possible approximation to objectivity if only we are not inclined towards self-deceit.

The attitude consisting in the rejection of every subjective element is only an involuntarily self-deceit as being not absolutely but only partially helpful and valid. Rejecting it, we can only lose something because what is objectively known will be at any rate considered, but it alone says nothing outside the realm of the certain consequences (if not arbitrary unjustified foresights), whilst adding some subjective judgment the way to probably forecasting is open.

The objectivistic position was once depicted by this analogy: "This ground is not sufficiently consistent: it is sand. Let us remove the sand, and ground the building on the void!" And Giuseppe Pompily (a friend moderately sympathetic with subjectivism, too early departed a few years ago) used to repeat, with reference to the need of implementing the objective data with subjective elements, such sentence by Pirandello: "A fact is like a sack: if it is void, it cannot stand upright".

Subjective elements are intended to fill this void, and seem necessary to this end. They will noways destroy the objective elements nor put them aside, but bring forth the implication that originate only after the conjunction of both objective and subjective elements at our disposal.

References[1]

Blyth, C. L. (1972). *The Amer. Statistician*, **26**, 20–22.
Boole, G. (1854). *The Laws of Thought* (repr. Dover publ.).
Bruno, A. (1964). Ch. 10 in de Finetti, 1972.
Cornfield J. (1967). *Review. Int. St. Inst.* **35**, 34–49.
De Morgan, A. (1847). *Formal Logic*. Taylor and Walten, London.
Dempster, A. P. (1972). *The Amer. Statistician*, **26**, 46.
de Finetti, B. (1970*a*). *Teoria d. Probabilità*. Einaudi, Torino.

[1] For a larger Bibliography see de Finetti, 1972; here are there only the references concerning specific quotations in the present text.

de Finetti (1970*b*). *Acta Psychologica*, **34**, 129–145.
de Finetti (1971). *St. in onore di G. Pompilj*. Oderisi, Gubbio.
de Finetti (1972). *Probability, Induction and Statistics*. Wiley, London.
Good, I. J. (1969). *J. Amer. Stat. Ass.* **64**, 23–66.
Kyburg, H. E., Smokler, H. E., eds. (1964). *Studies in Subjective Probability*. Wiley, New York.
Lindley, D. V. (1971). *Making Decisions*. Wiley Intersc., London.
Maritz, J. S. (1970). *Empirical Bayes Methods*. Methuen, London.
Mosteller, F., Wallace, D. L. (1963). *J. Amer. Stat. Ass.* **58**, 275–309.
von Mises, R. (1942). *Annals Math. Stat.* **13**, 156–165.
von Neumann, J., Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton Un., Pr. N.J.
Pearson, E. (1962). *Annals Math. Stat.* **33**, 394–403.
Polya, G. (1953). *Patterns of Plausible Reasoning* (Vol. II of "Mathematics and Plausible Reasoning"). Princeton Un., Princeton, N.J.
Pompilj, G. (1952). *Teoria dei campioni*. Veschi, Roma.
Pratt, J. W., (1961). *J. Amer. Stat. Ass.* **56**, 163–167.
Ramsey, F. P. (1926). *Truth and Probability* (repr. in Kyburg and Smokler, see).
Robbins, H. (1964). *Annals Math. Stat.* **35**, 1–20.
Roberts, H. V. (1965). *J. Amer. Math. Ass.* **60**, 50–62.
Savage, L. J. (1971). *J. Amer. Math. Ass.* **66**, 336, 783–801.
Suits, D. B. (1967). *Estudos Inst. Gulbenkian*, 231–289.
Tukey, J. W. (1962). *Annals Math. Stat.* **33**, 1–67.
Vaihinger, H. (1911). *Die Philosophie des "Als Ob"*, Berlin.
Wald, A. (1939). *Annals Math. Stat.* **10**, 299–326.

Résumé

Le rôle unifient de la théorie bayesienne pour les fondements et pour les applications de la statistique.

Afin qu'une discussion sur ce sujet puisse être féconde, il faut tout d'abord la dégager des difficultés qui surgissent des aspects particulièrs sont considérés séparément, comme des questions techniques distinguées concernant des problèmes isolés. L'induction est conçue d'ordinaire, en effet, comme un ensemble fragmentaire de théories très différentes, partielles et jamais satisfaisantes. Il y a la théorie formaliste d'un certain type de logiciens, et il y a les deux théories mathématiques qui s'adressent respectivement au «raisonnement inductif» et au «comportement inductif». Et tout devient de plus en plus compliqué à la suite d'une inflation de plus en plus lourde de formalisme: c'est le formalisme qui se présente comme une remède illusoire contre les insuffisances logiques causées par l'isolément, et qui agit par contre lui même comme facteur d'isolément.

La thèse developpée dans ce rapport est dirigée du côté exactement opposé, s'efforçant de montrer, aussi simplement et clairement que possible, que le sujet dans son intégralité admet une interprétation unique et tout-à-fait naturelle et une réponse simple et universelle, sous la seule condition d'éviter les préjugés diffusés dépendant de constructions artificielles et unilatérales. Cette simple réponse est celle de la théorie bayesienne; mais c'est dommage qu'on l'appelle «théorie» et qu'on lui a donné un nom: c'est le même regret exprimé par M. Cornfield qui fi dit – en faisant noter que le résultat de Bayes est tout-à-fait évident – «c'est exscessivement solennel de le qualifier de 'théorème'»!

La probabilité n'a qu'une seule signification réelle: celui de *dégré de confiance*: une signification subjective, bien sûr, même si chacun doit tenir compte reasonablement des données objectives connues, comme symmètries, fréquences, analogies, et tout autre expèrience. Et même s'il y a l'obligation de les évaluer de façon qu'elles soient cohéremment en accord entre elles et d'en faire évoluer cohéremment l'évaluation au fur et à mesure que se modifie l'étar d'information. C'est ceci – dit sommairement, que c'est le «théorème de Bayes».

Tout cela est imposé, et d'une façon unique, par effet de raisons concordantes provénant de chacun des aspects, mais qui ne forment, dans leur essence, qu'une seule raison dont seulement les apparences différent legèrement dans des situations différentes.

Ces thèses, que l'on ne pouvait ici qu'esquisser dans une forme abstraite et apodyctique, ont été par contre exposées dans le rapport avec tout le soin, en faisant la comparaison des consequences que l'on obtient en leur obéissant ou en s'en éloignant, par rapport à tout aspect des problèmes concernant l'induction, soit au point de vue logique ou du raisonnement et du comportement inductif.

Il faut repousser les objections contre la subjectivité: il faut remarquer en effet que, si l'on se passe des opinion subjectives, on n'obtient pas une plus grande objectivité; tout au contraire, une intégration subjective à l'ensemble stéril des données objectives résulte nécessaire dans le sens que, si l'on l'ignore, on est conduit à une situation où les décisions ne sont pas améliorées mais échappent à tout contrôle.