A GENERALIZATION OF BAYESIAN REASONING TO NON-PROBABILISTIC CASES

Jorge Carrera*

Abstract.

The success of Bayesian analysis has lead to an incredible production in statistics and probability theory but there have been much less efforts towards a generalization outside these disciplines. A new approach to the Bayesian scheme for a posteriori evaluation allows the construction of analogous schemesin other fields, where it could be as successful as in its probabilistic setting. First it is shown that a formal Bayesian scheme, presented under the viewpoint of system theory, can be translated to other fields. Examples in logic and graph theory show that Bayesian-like schemes function when the set of previous events or premises or nodes for the actual situation of the system is known. An evaluation of these events or premises is then calculated based on the previous information and on the characteristics of the system (probabilistic, logic, graph-theoretical, etc.). The dynamics of the systems is given via recursive implicit schemes for the step previous to the actual, z_{k-1}*. This is condensed in the definition of a new class of systems, retroactive systems, closely related to anticipatory systems, and specifically designed for applications analogous to the probabilistic setting but in other areas of applied mathematics.*

Keywords. Formal Bayesian scheme, non-probabilistic settings, backward systems, retroactive systems, decision's criteria.

Notation. All notations are standard.

* DIMEI, Facultad de Ingeniería, National University of Mexico, UNAM, 04510 México, D.F.
Senior Member IIAS

Introduction. A non-probabilistic generalization of Bayes' scheme Bayesian analysis is one of the main tools in probabilistic hypothesestesting, *a posteriori* analysis and its applications in many fields. This work is focused in this aspect of Bayesian analysis.

There are many specialization of the Bayesian methodology [see ISBA web-site] but not somuch attention has been given to generalizations outside probability theory. No references about this kind of systems were found. Thus, any advance in that direction seems to be new. A proposal is here presented.

It is shown that Bayesian and Bayesian-like systems can be seen as a realization of a wider class of systems that give information over previous steps of time(or previous values of some other parameter)as output.

On its turn, those systems build a subset of the "class of retroactive systems". It may well be seen as a kind of dual concept to the class of anticipatory systems proposed by Rosen and developed by [Dubois, 1998]. An anticipatory system is defined by a recursive function

$$x_{k+1} = F\left(\ldots, x_k, x_{k+1}, p\right) \quad (1)$$

where a "prediction" over the next time step, x_{k+1}, is made based on information about the actual time step, x_k, and the values of some parameters *p*. In the general case no complete information about the involved data in assumed and F does not needto be deterministic.

Under the same premises a *retroactive system* is here understood as a system whose dynamicsis given by a backward implicit scheme of the type

$$x_{k-1} = F\left(\ldots, x_{k-1}, x_k, p\right) \quad (2)$$

Bayesian and Bayesian-like systems are members of an important subclass of retroactive systems, the class of *retroevaluative systems*, those whose dynamics is given by a scheme of the type

$$G(x_{k-1}) = F\left(\ldots, Gx_{k-1}, Gx_k, p\right) \quad (3)$$

where *G* is an evaluative or a decision function, i. e. for all possible outputs of the system in a given step, *G* should give an evaluation parameter or some operativecriterion to choose an optimal state or output x_{k-1}.

The difference between general retroactive and retroevaluative systems is equivalent to the more technical difference between systems that construct or define the set of states or outputs that should have lead to the present situation; and the case where the forerunners are known and a decision has to be made as to what state or output is to be chosen, if possible with a measureof the error or "cost" of the cases when the chosen value is only approximately the exact value. The concepts here presented are closely involved with optimization theory.

General retroactive systems are not treated here. They are to be developed in future papers. It is shown how the main formal features ofBayesian-type systems localized and abstracted in a process of three representative examples.

The first one is the well-known Bayesian scheme. The others are presented in a similar notation in order to emphasize the similarities.

The second example is a logical analogy of the Bayesian case. This example is the most elaborated because it shows how an analogy to Bayesian systems is constructed formultivalued logical systems. At the same time, in contrast to the probabilistic independence of the events considered in the Bayesian case, a logical formula,as a step in a deduction chain, will be in general derived from more than one formula. It is shown that the logical scheme does not necessarily require a condition analogous to the events' independence in probability. In this sense no "superposition principle" (the only interaction between possible "causes" is through addition) can be assumed and the system could be designed as "non-linear". See Note 05

The example of graph theory shows some "bare" structure of Bayesian-type systems and again the possibility to have more than one solution.

Case 04 is a plausible proposal, not yet developed, of the application of the methodology to a field far away from the previous examples, control theory.

Bayesian analysis is very fruitful and continues to develop in many fields of application. It is then to be expected that this kind of analogous reasoning can be as successful as its "classical" probabilistic paradigm.

Part 1. Four representative cases

Case 01. *Bayesian systems*

Let B be an event in a well-defined probabilistic space with a known probability P(B). Assume there is a set of n independent events A_k such that

$$P(B!A_1 \cup A_2 \cup \ldots \cup A_n) = 1. \qquad (4)$$

This implies

$$P(B) = \Sigma P(BA_i) = \Sigma P(A_i)P(A_i|B) \quad (5)$$

Under these premises the problem is:

Problem 01. Find the conditional probability of A_k once it is known that B is given:

$$P(A_k|B) = \gamma_k \quad (6)$$

Data. Probability of B, P(B),

Probabilities of the A_k:

$$P(A_k) = \alpha_i$$

Conditional probabilities of B given A_k:

$$P(B|A_k) = \beta_k$$

Solution. It is given by Bayes formulae:

$$P(A_k | B) = \frac{P(A_k)P(B|A_k)}{P(B)} \quad (7)$$

$$P(A_k | B) = \frac{P(A_k)P(B|A_k)}{\sum_j P(A_j)P(B|A_j)}$$

Note 01. These formulae have a natural backward recursive form.

Case 02. *A logical case*

Let L continuous valued first-order logic over the "classic" set of well-formed formulae (wff) of first order logic is given. Much of the work here presented was made using the original proposal of Nilsson of a probabilistic logic [Nilsson, 1986]. Working with that logic gave insight and allowed the author to generalize over it. Other multivalued logics also do the job, as long as the following statement is true, as an axiom or theorem:

Statement 01. Be T the function assigning a truth-value $\alpha \in [0, 1]$ to a well-formed formula (wff) p. Then

$$T(\neg p) = 1 - T(p), \quad (8)$$

Where $\neg p$ is the negation of p.

Denote by B a wff that is not an axiom. Consider the set of all syntactic deductions of B from the axioms, without repetitions. The identity deduction $\frac{p}{p}$ (given p, p is deduced) is considered identical to just *p*.

In general, for a given expression B there will be more than one deduction and there will be more than one wff in the deductive step previous to B, the one before last in every deductive chain ending in B. Assume first there is only one formula in the previous step to B in every deductive chain leading to B. Call these formulae *immediate predecessors* of B and denote them by A_{k-1}, The truth value of the A_{k-1}, $T(A_{k-1})$ is known as well as the truth-values

$$T(A_{k-1} \rightarrow B) \qquad (9)$$

of the implications $A_{k-1} \rightarrow B$.

Problem 02. Find the truth-value of the implications

$$B \rightarrow A_{k-1},\ T(B \rightarrow A_{k-1}) = \gamma_k, \qquad (10)$$

under the premise that

$$T(A_{k-1,1} \vee A_{k-1,2} \vee \ldots \vee A_{k-1,n} \rightarrow B) = 1. \qquad (11)$$

Note 02. This problem is not unambiguous.Just as B can have many antecessors, every A_{k-1} can have more than one successor. This is one of the reasons why the choice of a multivalued logic is adequate.

Note 03. Depending on the assumptions made and on the type of logical structure, there can be more than one possibility to calculate a possible value for the implications in (10). Here a solution very close to the Bayesian formula is given.

Solution. It is given by

$$T(B \rightarrow A_k) = T(B) - T(A_k) + T(A_k \rightarrow B) \qquad (12)$$

The formula is valid for the classical bivaluated case, as can be calculated giving truth-values. In a more general case it can be deduced in the following way:

$A_k \rightarrow B$

$\neg A_k \vee B$

$\neg\neg(\neg A_k \vee B)$

$\neg(\neg(\neg A_k \vee B))$

$\neg(\neg\neg A_k \wedge \neg B)$

$\neg(A_k \wedge \neg B)$

Taking truth-values

$T(A_k \rightarrow B) =$

$T(\neg(A_k \wedge \neg B))$

$1 - T(A_k \wedge \neg B)$

$1 - T(A_k)T(\neg B)$

$1 - T(A_k)(1 - T(B))$

$1 - T(A_k) + T(A_k)T(B)$ (13) Doing the same with the formula $B \rightarrow A_k$ one gets

$$T(B \rightarrow A_k) = 1 - T(B) + T(B)T(A_k) \quad (14)$$

Putting (12) and (13) as

$$T(A_k)T(B) = T(A_k \rightarrow B) + T(A_k) - 1 \quad (15)$$

$$T(B)T(A_k) = T(B \rightarrow A_k) + T(B) - 1 \quad (16)$$

Both right sides are then equal

$$T(B \rightarrow A_k) + T(B) - 1 = T(A_k \rightarrow B) + T(A_k) - 1 \quad (17)$$

From which (12) is obtained. This deduction is based on the reasoning of the basic deduction of Bayes formulae (7).

Note 04. Formula (12) has also a natural backwards recursive form.

Note 05. It was first assumed that B has only one previous formula as immediate predecessor in its deductive chain. However, this assumption is more restrictive in this case than the independence of events in the probabilistic one. More often than not a formula is derived from two or more formulae. However, condition (11) does not imply with necessity that Bshould be derived only from one A_{k-1}. Also, the assignation of a truth-value less than 1 to the statement A_{k-1} can be interpreted as the failure of A_{k-1} to be a formula from which B can be derived without other formulae. In this sense formula (12) does not lose its validity if A_{k-1} is not the only formula from which B is derived.

But if, in opposition to the classical probabilistic case, “events” (in this case formulae) can interact, no superposition principle applies and thus the problem is “non-linear”.

Case 03. *Graph theory*

Let G be a “sufficiently big” directed and weighted graph. To analyze the local internal structure of G assume B is a node at least two steps (one step means one arrow) away from any initial or end node, if there are. Consider the set of all nodes $A_{k-1,j}$one-step"before" B. That is the set of nodes separated by one arrow from B, directed from $A_{k-1,j}$ to B, j = 1,… ,n

Let the weights $W(Ak,B)$ of the directed arrow A_kB be given, under the condition that

$$\sum_{k=1}^{n} W(A_k B) = 1 \tag{18}$$

Problem 03. Once node B has been reached, calculate weights for the inverse arrows BA_k.

Solution. In this generality the problem is almost just a formal scheme and certainly more than one solution can be given, if a solution exists. Thus the question arises: under which premises is it possible to find or construct only one function F that allows the calculations of the sought weights using information analogous to the previous cases?

Once the examples has been analyzed, an answer can be given: If F is a function such the following scheme is recursive and backwards implicit and defined for the sought information A_{k-1} "previous" to B,

$$W(A_{k-1}) = F(\ldots, W(A_{k-1}), W(B), p)) \tag{19}$$

then the scheme gives correctly defined weights for the arrows BA_k.

Note 06. Introduction to retroactive systems. Just like the schematic representation of anticipatory systems is given by a state- or output-function F such that a recursive forward implicit scheme

$$x_{k+1} = F(\ldots, x_k, x_{k+1}, p) \tag{20}$$

is well defined for some parameters *p*, a purely *retroactive* (non-evaluative) system is given by a recursive backward implicit scheme of the form

$$z_{k-1} = G(z_{k-1}, z_k, \ldots, p) \tag{21}$$

and a *retroactive evaluative* system is given by a function

W: Set of states or set of outputs $\rightarrow$ [0, 1]

That defines a recursive implicit backward scheme

$$W(z_{k-1}) = G(W(z_{k-1}), W(z_k), \ldots, p)) \tag{22}$$

Note 07. The choice of the interval [0, 1] is conventional. Other connected ordered intervals or subsets can be used.

The functions W, F and G can satisfy some optimality condition.

Examples of evaluative functions are the assignation of probabilities in the Bayesian case and the truth-value functions.

Case 04. *Control theory*. The problem of attainability is the determination of the set of states of a system that can be "controlled" to a desired state B. It has two aspects:

a) Given a state u_0 establish if it belongs to the set of attainable states or not, and
b) Construct or define the set of attainable states.

A typical control problem is the stability of a process. If some of the parameters take values beyond certain limits, the state of the process changes to "nearing malfunctioning" and an action is taken to return the system (control it) to the desired set of states. If the process is sufficiently complex, it is not always clear which part of the system failed, information that is crucial to the control of the process. This can be done with the classical Bayes scheme in terms of probabilities. But this can also be done in some other terms. A fuzzy or neuronal control is certainly an adequate structure to define systems like the ones here presented.

Part 2. Evaluative retroactive systems(retroevaluative systems)

Definition 01. Let S = (I, O, Z, f, g) be a system, where I is the set of inputs, O the set of outputs, Z the set of states, f the state transition function

$$f: Z \times I \to Z,\ f(z_1, i) = z_2,\quad z_1, z_2 \in Z \qquad (23)$$

and g the output function

$$g: I \times Z \to O$$

S is called an evaluative retroactive system If there is a function W of Z into a well-defined subset of an ordered set M, called *evaluation function*that defines an implicit backward recursive scheme for $W(x_{k-1})$

$$W(x_{k-1}) = F(W(x_{k-1}), W(x_k), \ldots, p), \qquad (24)$$

Where F is a function of the W-values of the present state, the previous state and other possible parameters*p*, like the values of all possible previous steps leading to x_k besides x_{k-1}, if there are. The requiered W-values can be given at each step, as measures, approximations or estimations, or they are to be constructed assuming that the function giving the state x_{k-1}

$$x_{k-1} = G(x_{k-1}, x_k, \ldots)), \qquad (25)$$

defines a retroactive system.

The same is valid when the codomain of G is the output set O.

Definition 02. S is a *retroactive system* if either its state-transition function or its output function defines a system of the form (25).

Note 07. In the Bayesian case it is known that if a function $C(x_E, x_0)$ is defined, giving some measure of the "cost" of event x_E not being the "real" event x_0 when the decision is made, the expected value or first moment of C, E(C), is called *probabilistic cost function*. If the decision is based on a Bayesian criterion, then the corresponding cost function is minimized.

The truth-value assignation of Case 02 is calculated based on internal properties of the considered multivalued logic. In some cases W gives only the values of a parameter with values on a ordered set M. This order allows the definition of some optimality criteria or function.

Note 08. Evaluative retroactive systems can also be seen as general retroactive systems.

Part 03. Retroactive systems

This concept was developed because of the necessity to generalize the efficient Bayesian analysis to situations where a probability setting is not involved or not desired. On the other hand, retroactive systems are a kind of "dual" to anticipatory ones.

Dubois [1, p. 4] defines "an incursion (as) an inclusive or implicit recursion... in the following way

$$x(t+1) = f[x(t), x(t+1); p)]" \quad (26)$$

with some parameter p.

This "defines a self-referential system which is an anticipatory system of itself… and it "contains [substituting (24) in x(t+1)] a model of itself".

Certainly, Dubois is defining a very wide class of systems.He leaves the concept of a "model of itself" in some needed ambiguity. In this study the concept of "model" is assumed as a concrete realization of a theoretical retroactive system defined by

$$S = (I, O, Z, f, g)$$

where f or g define retroactive schemes.

As a "concrete realization" of S it is understood that a set of finite data is measured, approximated, calculated or estimated that represents the corresponding values of the parameters and other data defining a state of a subset of a"real" system from which it is assumed that its laws are represented by the functioning of the system. Once these data are feed into the system, its output can be again made to correspond, with some established precision, to other measured, approximated, calculated or estimated data.

Retroactive systems, in the form here presented, are focused in applications, especially where Bayesian-like, but not probabilistic, thinking is required.

Part 04. Applications

It has been schematically shown that there are logical systems that do not need to use a probabilistic interpretation to be examples of retroactive evaluative systems, as well as some graphs.

A "fuzzy control theory" gives natural candidates for retroactive systems and subsystems.

As is also the case with anticipatory systems with the corresponding forward-system, backward implicit finite-differences deterministic schemes are a good model of the scheme here presentedand it is possible to construct or analyze non-deterministic schemes.

In general, a new field of application tools is opened here:*generalized non-probabilistic Bayesian-type schemes*. The discussion shows that in its broader setting these systems can be seen as part of a theory which is "dual" to a already known and developed theory, anticipatory systems, from which experiences and applications can be derived. Optimality criteria are also essential part of the schemes here presented.

Part 05. Conclusions

Bayesian analysis is one of the most successful tools in many areas of science. It has been well defined for many different cases and many different and intelligent applications have been made [2].

But generalizations of other types, especially outside statistics and probability, have not attracted attention. Here a generalization of the formal aspects of the basic Bayesian scheme is presented and a structure developed that can be re-constructed in non-probabilisticareas.

These systems turned out to build a kind of dual concept to anticipatory systems, although they have some intrinsic properties due to its direct relation to certain applications and its backward nature. A wide window of possibilities is opened.

Bibliography

[1] Dubois, D. M., *Introduction to Computing Anticipatory Systems*, International Journal of Computing Anticipatory Systems, CHAOS, vol. 2, pp. 3-14, 1998.

[2] SBA, Int. Society for Bayesian Analysis, Bayesian.org

[3] Nilsson, N. J., *Probabilistic logic*, Artificial Intelligence, Nr. 28, Elsevier, 1986.