Economic Notes by Banca Monte dei Paschi di Siena SpA, vol. 26, no. 3-1997, pp. 477-486

A Note on the Subtleties of Bayesian Inference

GIACOMO BONANNO*

Bayesian theory asserts that the beliefs formed by a rational individual in response to new information must be derived from the original beliefs by conditioning on the information. Information is represented by a set of possible states. A state, in turn, ought to represent a complete description of the world. Once the states have been described accurately, application of Bayes' rule is a non-problematic and mechanical procedure. However, if the states have not been described fully and correctly, then inference based on Bayes' rule will typically be incorrect. We illustrate this with the help of a case discussed by Nalebuff.

(J.E.L.: C72).

The Bayesian approach plays a central role in economics, decision theory and game theory. Bayesianism is usually characterized as the philosophical view that probability can be interpreted subjectively[1] and that the rational way to assimilate information into one's structure of beliefs is by a process called "conditionalization"[2]. Thus Bayesianism has a static part and a dynamic part. The former asserts that a coherent set of beliefs can be represented by a probability function over sentences or events (see De Finetti, 1937, Ramsey, 1931, Savage, 1954, and, for a recent survey, Hammond, in press). The dynamic part of Bayesian theory asserts that the beliefs formed by a rational individual in response to new information must be obtained from the original beliefs by conditioning on the information. That is, if the individual starts with a subjective

* Department of Economics, University of California, Davis, CA 95616-8578. U.S.A. e-mail: gfbonanno@ucdavis.edu

I am grateful to Klaus Nehring and two anonymous referees for helpful comments and to Ignacio Ortuño-Ortin for bringing Nalebuff's contribution to my attention.

[1] Although there are also Bayesians who do not subscribe to the subjectivist interpretation of probability.

[2] The qualitative implications for belief revision of the rule of conditionalization are studied in Battigalli and Bonanno (1997).

probability distribution P_0 and learns E, where $P_0(E) > 0$, then her new beliefs should be given by the probability distribution P_n defined as follows: for every event A, $P_n(A) = \frac{P_0(A \cap E)}{P_0(E)}$. The conditionalization rule thus says that the correct way to accommodate a piece of information into a belief set is by revising the initial subjective probabilities according to Bayes' rule. We refer to this as "Bayesian inference". There is an ongoing debate in the philosophical literature as to whether Bayesian inference is a requirement of rationality (see, for example, Brown, 1976, Howson and Urbach, 1989, Jeffrey, 1965, Maher, 1993 and Teller, 1973). The purpose of this note is not to contribute to this debate but to highlight the subtleties of Bayesian inference. In particular, of crucial importance is the interpretation of an event. Information is usually represented as an event, that is a set of states. A state, in turn, ought to be thought of as a *complete description of the world.* Once the states have been described accurately, application of Bayes' rule is a non-problematic and mechanical procedure. However, if the states have not been described fully and correctly, then inference based on Bayes' rule will typically be incorrect. We illustrate this point with the help of a real world problem discussed by Nalebuff (1990).

A 42-year-old woman died suddenly a few hours following the extraction of a wisdom tooth. The coroner's report showed that a possible cause seemed to be an allergic reaction (anaphylactic shock) to a drug. The dentist had prescribed a pain medication, zomepirac, which she was to take if needed. It is not known whether or not she took it, however the *ex ante* probability of her taking the drug was 0.6[3]. Furthermore, we are told that, conditional on the information that a patient takes zomepirac and dies, the probability that the drug is the cause of death is 0.95. Finally, we can also assume that the probability of a patient dying of some cause unrelated to zomepirac is the same, whether or not the patient takes the drug. We are asked to compute the probability that zomepirac was the cause of the woman's death. A tempting answer is (0.60) (0.95) = 0.57 (the chance she took it times the chance it killed her). Nalebuff rejects this answer on the grounds that it ignores an important piece of information, namely the fact that the woman did die. He suggests the following as the correct inference. There are 5 possibilities as described in the following table (the symbol ^ stands for "and"):

[3] A study of other patients who had similar operations shows that 60% of them had sufficient post-operative pain that they took the pain medication prescribed.

Table 1

possibility	1	2	3	4	5
description	(patient **takes Z)** ^ (patient **dies because of Z)**	(patient **takes Z)** ^ (patient **dies of unrelated causes)**	(patient **takes Z)** ^ (patient does **not die)**	(patient **does not take Z)** ^ (patient **dies)** [hence of unrelated causes]	(patient **does not take Z)** ^ (patient does **not die)**
probability	x_1	x_2	x_3	x_4	x_5

Thus, for example, the event "the patient takes the drug" is represented by the set of possibilities {1, 2, 3}. The above assumptions are translated into the following four equations:

$x_1 + x_2 + x_3 + x_4 + x_5 = 1$ (the x_i's are probabilities)

$x_1 + x_2 + x_3 = 0.6$ (the *ex ante* probability of taking the drug is 0.6: cf. footnote 3)

$\frac{x_1}{x_1 + x_2} = 0.95$ (conditional on the information that a patient takes zomepirac and dies, the probability that the drug is the cause of death is 0.95)

$\frac{x_2}{x_1 + x_2 + x_3} = \frac{x_4}{x_4 + x_5}$ (the probability of a patient dying of some cause unrelated to zomepirac is the same, whether or not the patient takes the drug)

Although these equations are not sufficient to solve for the probabilities, they *are* sufficient to answer the question. In fact, the above system of equations allows us to express each variable in terms of x_4, in particular, $x_1 = 28.5\ x_4$ and $x_2 = 1.5\ x_4$. Using this, we can compute the probability that the patient died because of the medication, given the information that the patient did die, as follows:

$$\frac{x_1}{x_1 + x_2 + x_4} = \frac{28.5\,x_4}{28.5\,x_4 + 1.5\,x_4 + x_4} = \frac{28.5}{31} = 0.919$$

Thus the conclusion we are offered is that the probability that the drug was the cause of the woman's death is as high as 92% (much higher than the "naïve" estimate of 57%).

We will argue, however, that the above analysis is based on an incomplete "description of the world". Once a complete analysis is carried out it turns out that the above conclusion is either incorrect or is based on a rather strong implicit assumption, which seems hard to justify.

As Savage (1954, p. 9) pointed out, a possibility, or state, should be a "complete description of the world, leaving no relevant aspect undescribed". A state is thus described in terms of a number of propositions that are considered relevant to the analysis. In our example, two such propositions are clear: "the patient **dies**" (call it **D** and its negation ¬D) and "the patient **takes** zomepirac" (call it **T** and its negation ¬T). A third proposition is suggested by the coroner's report, namely "the patient is **allergic** to zomepirac" (call it **A** and its negation ¬A). Finally, there is the proposition "the patient dies of other causes". To make the analysis more concrete, suppose that "other causes" means a deadly virus. Then we denote by **V** the proposition "the patient has the **virus**" and by ¬V its negation. The "possible worlds" are represented by the end nodes of the tree in Figure 1. Greek letters denote (subjective or objective) probabilities. Ignore for the moment the dotted lines and the extra two nodes marked VATDz and VATDv. The following, very plausible, assumptions are reflected in the figure: (1) the probability that the patient

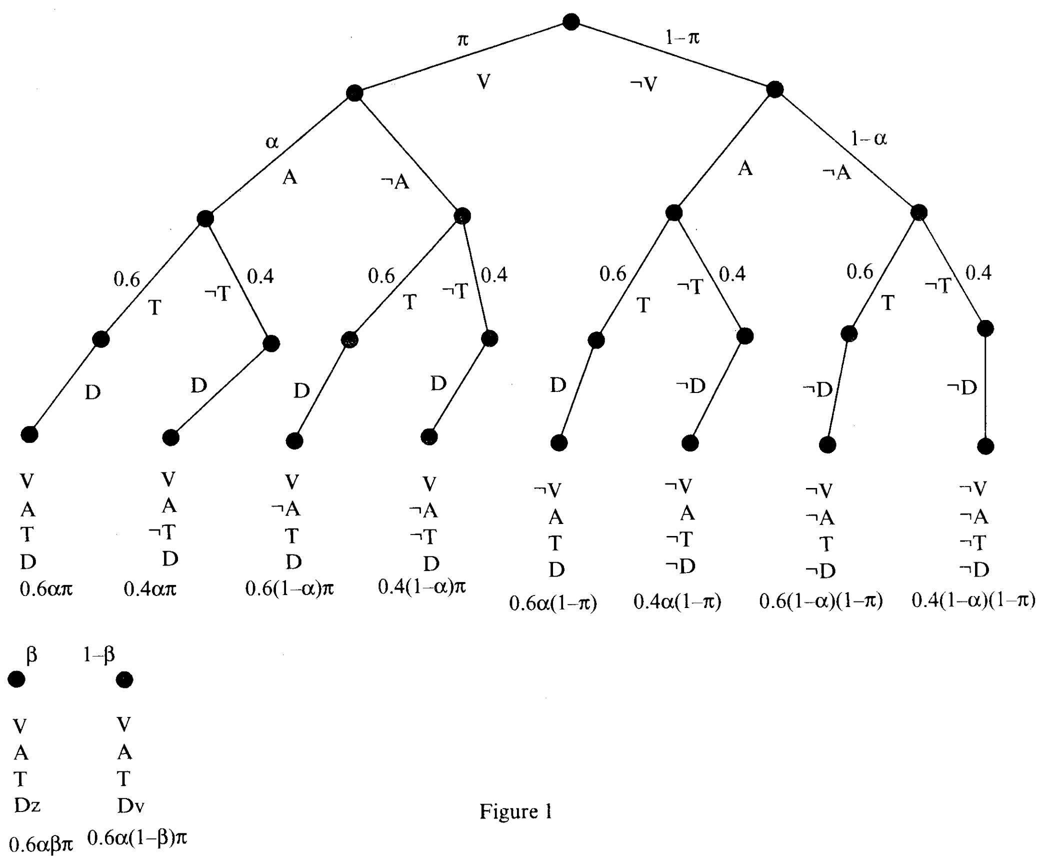

Figure 1

takes zomepirac (= 0.6) is the same, whether or not she is allergic to it (because she does not know: if she knew she would take it with probability zero!) and is also not affected by her having or not having the virus; (2) the virus is a sufficient cause of death, as is an allergic reaction to the drug; (3) there is no other cause of death; (4) the probability of having the virus (denoted by p) does not depend on whether the patient is allergic to zomepirac or not and, conversely, the probability of being allergic to zomepirac (denoted by α) does not depend on whether or not the patient has the deadly virus.

The crucial step now is to identify the event that represents the proposition "the patient dies because of the drug". We shall discuss alternative interpretations and argue that none of them justifies the conclusion that the probability that the woman died because of the drug is 92%.

First interpretation (stressing the cause-effect relationship). The proposition "the patient dies because of the drug" is clearly true at world ¬VATD. But what about world VATD? Here we have two factors (the virus and the allergic reaction to the drug), each sufficient to cause death. Thus whether the actual death is caused by one factor or another depends on which of the two reactions (to the virus or to the drug) took place first and we are not given this information. Let us therefore denote by Dz the proposition "the patient dies because of zomepirac" (i. e. the allergic reaction kills before the virus) and by Dv the proposition "the patient dies because of the virus " (i. e. the virus kills before the allergic reaction takes place). Let β be the probability of the former and $(1-\beta)$ the probability of the latter. These two descriptions of the world are denoted by dotted lines in Figure 1. We are now in a position to interpret the sentence "the patient dies because of the drug": the sentence is represented by the event

$$B = \{VATDz, \neg VATD\}.$$

From Figure 1 we get

$$Prob(B) = 0.6\,\alpha\,(\beta\,\pi + 1 - \pi).$$

Let Δ denote the event "the patient dies" and Z the event "the patient takes the drug and dies".

Then

$$\Delta = \{VATDz, VATDv, VA\neg TD, V\neg ATD, V\neg A\neg TD, \neg VATD\} \text{ and}$$

$$Z = \{VATDz, VATDv, V\neg ATD, \neg VATD\}.$$

It is easy to see from the figure that

$$\text{Prob}(\Delta) = \pi + 0.6\,\alpha\,(1-\pi) \quad \text{and} \quad \text{Prob}(Z) = 0.6\,\pi + 0.6\,\alpha\,(1-\pi).$$

We have to determine

$$\text{(1)} \qquad \text{Prob}(B\,|\,\Delta) = \frac{\text{Prob}\,(B \cap \Delta)}{\text{Prob}\,(\Delta)} = \frac{\text{Prob}\,(B)}{\text{Prob}\,(\Delta)} = \frac{0.6\,\alpha\,(\beta\pi + 1 - \pi)}{\pi + 0.6\alpha\,(1 - \pi)}$$

The extra information we are given is that, conditional on taking the drug and dying, the probability that death was caused by the drug is 95%:

$$\text{(2)} \quad \text{Prob}(B\,|\,Z) = \frac{\text{Prob}\,(B \cap Z)}{\text{Prob}\,(Z)} = \frac{\text{Prob}\,(B)}{\text{Prob}\,(Z)} = \frac{0.6\,\alpha\,(\beta\pi + 1 - \pi)}{0.6\pi + 0.6\alpha\,(1 - \pi)} = 0.95$$

Notice first that (2) implies that $\pi > 0$, that is, the probability that the woman had a deadly virus cannot be zero.

Solving (2) for α we get $\alpha = \dfrac{19\pi}{1 - \pi + 20\beta\pi}$. Substituting this into (1) we get

$$\text{(3)} \qquad \text{Prob}(B\,|\,\Delta) = \frac{28.5\,(1-\pi) + 28.5\,\beta\pi}{31(1-\pi) + 50\,\beta\pi}$$

Now, from (3) we have that, since $\pi > 0$ [it follows from (2)], $\text{Prob}(B\,|\,\Delta) = \frac{28.5}{31} = 0.919$ *if and only if* $\beta = 0$. Thus the analysis suggested by Nalebuff implicitly assumes that $\beta = 0$, that is, that if the woman died when both factors were present (the virus and the allergic reaction to the drug) then *death must be ascribed to the virus.* One can consistently assume this, but it does seem to be a rather strong assumption that is not justified by the data of the problem.

Furthermore, even if this assumption is considered plausible, we were not even made aware of it in the original discussion. When $\beta > 0$, (3) can take *any* values between 0.57 and 0.919. Note also that for every $\pi > 0$ and $\beta > 0$, Prob $(B\,|\,\Delta)$ is strictly *less* than 0.919. Given the lack of information as to which of the two factors can indeed be considered the true cause of death, when both factors are present (state VATD), a reasonable assumption is that they are equally likely as the actual cause of death. In the above framework this corresponds to $\beta = \frac{1}{2}$. Substituting $\beta = \frac{1}{2}$ into (3) we obtain

$$\text{(4)} \qquad \text{Prob}(\text{B} \mid \Delta) = \frac{28.5 - 14.25\,\pi}{31 - 6\,\pi}$$

which is a strictly decreasing function. As $\pi \to 0$ (recall that, by (2), $\pi > 0$), (4) tends to 0.919 and as $\pi \to 1$ (4) tends to 0.57 ($= 0.95 \times 0.6$), the "naive" answer [from (3) we have that another case where $\text{Prob}(\text{B} \mid \Delta) = 0.57$ is when $\pi = 1$, for arbitrary β; in fact, if $\pi = 1$ then it follows from (2) that $\alpha\beta = 0.95$].

In conclusion, in this first interpretation of the proposition "the woman died because of the drug" we get that *every probability between 0.919 and 0.57 can in fact be justified!*

Second interpretation (a point of semantics). One could try to defend Nalebuff's analysis by reducing the matter to a point of semantics as follows: in the case where both factors are present (state VATD), one cannot say that the patient dies *because of the drug*, since she would die anyway, due to the virus. Hence setting $\beta = 0$ is justifiable. However, the same interpretation of the expression "the patient dies because of" leads to the conclusion that, when both factors are present, it is also *not* true that the patient dies because of the virus (she would die anyway, due to the allergic reaction to the drug). Hence setting $1 - \beta = 0$ is also justified, yielding a contradiction. That is, if the claim that she did not die because of the drug relies on the argument that she would have died anyway (when both factors are present), then the same argument yields the claim that she did not die because of another cause, contradicting the premise that "either the woman died because of the drug or she died because of another cause" (recall that, in our analysis, the virus represents "a drug-unrelated factor sufficient to cause death", that is, "another cause").

Third interpretation (the drug as a potential cause of death). Finally, one could interpret the proposition "the woman died because of the drug" loosely in the sense that perhaps the drug was not the actual cause of death but it could have been, that is, the drug either did or could have caused the death. With this interpretation, the proposition is identified with the event that the woman was allergic to the drug and took it. We call this event P. Thus

$$\text{P} = \{\text{VATDz}, \text{VATDv}, \neg\text{VATD}\}$$

Then $\text{Prob(P)} = 0.6\alpha$ (the probability that she took the drug times the probability that she was allergic to it). We have to determine the probability that the drug was a potential cause of death (event P) given the information that she died (event Δ):

$$\text{(5)} \qquad \text{Prob}(P \mid \Delta) = \frac{\text{Prob}\,(P \cap \Delta)}{\text{Prob}\,(\Delta)} = \frac{\text{Prob}\,(P)}{\text{Prob}\,(\Delta)} = \frac{0.6\,\alpha}{\pi + 0.6\alpha\,(1-\pi)}$$

We are given the following estimate:

$$\text{(6)} \quad \text{Prob}(P \mid Z) = \frac{\text{Prob}\,(P \cap Z)}{\text{Prob}\,(Z)} = \frac{\text{Prob}\,(P)}{\text{Prob}\,(Z)} = \frac{0.6\,\alpha}{0.6\pi + 0.6\alpha\,(1-\pi)} = 0.95$$

that is, if we knew that the patient died having taken the drug then we would attach probability 95% to the fact that the drug was a potential cause of death. Note again that (6) implies that $\pi > 0$. Solving (6) for π we get

$$\text{(7)} \qquad \pi = \frac{\alpha}{19(1-\alpha)}$$

which increases from 0 to 1 as α increases from 0 to 0.95. Thus from (7) we get that $\alpha \leq 0.95$; furthermore, since, by (6), $\pi > 0$, it follows that $\alpha > 0$. Substituting (7) into (5) we obtain

$$\text{(8)} \qquad \text{Prob}(P \mid \Delta) = \frac{28.5(1-\alpha)}{31-30\alpha} \qquad (0 < \alpha \leq 0.95)$$

Now, as α increases from 0 to 0.95, (8) decreases from 0.919 to 0.57. Once again, every value between the answer suggested by Nalebuff (92%) and the "naïve" answer (57%) is possible. Furthermore, the value of 92% is only possible in the limit as α tends to zero: for every admissible value of α, $\text{Prob}(P \mid \Delta) < 0.919$.

REFERENCES

P. BATTIGALLI - G. BONANNO (1997), "The Logic of Belief Persistence", *Economics and Philisophy*, 13, pp. 39-59.

P. M. BROWN (1976), "Conditionalization and Expected Utility", *Philosophy of Sciences*, 43, pp. 415-19.

B. DE FINETTI (1937), "Le prévision: ses lois logiques, ses sources subjectives", *Annales de l'Institut Henri Poincaré*, 7, pp. 1-68.

P. HAMMOND (in press), "Subjective Expected Utility and Non-cooperative Games", in S. Barberà, P. Hammond and C. Seidl (eds), *Handbook of Utility Theory*, Kluwer Academic Press.

C. HOWSON - P. URBACH (1989), *Scientific Reasoning*, La Salle, III, Open Court.

R. JEFFREY (1965), *The Logic of Decision*, McGraw-Hill, New York.

P. MAHER (1993), *Betting on Theories*, Cambridge University Press, Cambridge.

B. NALEBUFF (1990), "The Bayesian Butler Did It", *Journal of Economic Perspectives*, pp. 180-1.

F. RAMSEY (1931), *Foundations of Mathematics and Other Logical Essays*, Routledge and Kegan Paul, New York.

L. SAVAGE (1954), *The Foundations of Statistics*, John Wiley, New York.

P. TELLER (1973), "Conditionalization and Observation", *Synthese*, 26, pp. 218-58.