FREQUENCY-DRIVEN PROBABILITIES IN QUANTITATIVE CAUSAL ANALYSIS

Federica Russo
Draft of 9th November 2006

Abstract

This paper addresses the problem of the interpretation of probability in quantitative causal analysis. I argue that probability has to be interpreted according to a frequency-driven approach because in this way we can account for the different use and meaning of probability in generic and single-case causal inferences involved in this domain.

Contents

1. Introduction

A large part of the social sciences, e.g. demography, economics, sociology or epidemiology, aims at establishing causal relations. A long tradition initiated with the pioneering works of Adolf Quetelet (1869) in demography, Emile Durkheim (1897) in sociology and Sewall Wright (1934) in population genetics sees statistical models as a useful device to achieve this goal. As far as causal inference is concerned, two categories ought to be distinguished: generic and single-case. Whilst the former have in view population-level causal relations, the latter targets particular individuals; yet, the two share an important feature: they are both probabilistic. Whence the question: how is probability to be interpreted? This question is particularly relevant for two different but related reasons. Firstly, generic and single-case statements bear different meaning of probability and therefore call for different interpretations of probability. Secondly, single-case causal inferences are often used for making decisions and different interpretations can yield different probability values. I shall defend the view that within a Bayesian framework we'd better off with a *frequency-driven* approach, which embraces both the empirically-based and the objective Bayesian interpretation of probability. The argument will run as follows. After having presented quantitative causal analysis and the possible interpretations of probability within the Bayesian framework, I distinguish two types of inferences: generic and single-case. I then show that the meaning of probability is different therein: it is related to frequency of occurrence in the first, and to credence in the second. I defend the plausibility of a dual concept of probability and particularly the view that degrees of belief

need to be frequency-driven. I finally argue that the empirically-based or objective Bayesian approaches give a coherent interpretation that suits well the case at the generic and at the single-case level.

2. Probabilistic causal inferences in quantitative causal analysis

A panoply of disciplines falls under the label "social sciences". These disciplines investigate society from different angles and perspectives sometimes with radically different methodologies. Yet they all share the common objective of understanding, predicting and intervening on society and/or on individuals. For instance, demographic studies are interested in variations that occur in populations due to mortality, fertility and migration behaviour; economics studies the management of goods and services. Arguably, epidemiology also belongs to the social sciences insofar it is interested in the distribution of disease across populations. Knowledge of causes proves to be a necessary element both for the cognitive goal - i.e. understanding the causal mechanisms - and for the action-oriented goal - i.e. predicting and intervening at the population level as well as at the individual level on the basis of knowledge of those mechanisms.

In this paper, I shall focus on *quantitative* causal analysis, that is causal analysis performed by means of statistical models, e.g. structural models, covariance structure models or multilevel analysis. In the following, I shall take structural models as a paradigmatic example of different types of statistical models. A simple form of a structural equation is the following:

$$Y = \beta X + \varepsilon$$

where Y is the response variable, X is the explanatory variable, β is a parameter and ε represents the errors. An explicit causal interpretation takes Y to be the effect and X the cause. The equation states that a unit change in X will correspond to a β-unit change in Y. In other words, the structural equation attempts to determine the variation in Y (the effect) due to X (the cause). Probability comes in through different interpretations of the error term ε. Errors can be interpreted as lack of knowledge or as genuine chancy elements. The first option conveys the idea that causal relations are deterministic and that probabilities are introduced because of our partial and incomplete knowledge about the world. Alternatively, this first option can convey the idea that the world is stochastically represented, but this is not necessarily an ontological commitment to indeterminism. The second option, instead, says that the world *is* genuinely indeterministic and errors thus represent that part of the relation escaping our understanding and control.

Then, it seems that quantitative causal analysis forces us to make up our mind about determinism: is causality deterministic or indeterministic? Quite to the contrary, we are not obliged to push the discussion thus far. Whether or not Nature is governed by deterministic laws falls beyond the scope of the present discussion. The point at stake is that structural equations represent *probabilistic* causal relations and that, consequently, probabilities entering structural equations are to be *interpreted*.

The issue of the interpretation of probability is particularly relevant when *causal* inference is involved. Consider for instance epidemiological studies about the effects of tobacco consumption on lung cancer. For policy reasons (among

others), we are interested in ascertaining what the incidence of tobacco on lung cancer is, namely we are interested in a *generic causal relation* that holds at the population level. However, we might also be interested in *single-case causal relations* - for instance, in Harry's chance of developing lung cancer, given that he smokes, or in the probability that smoking actually caused him to develop cancer. At the generic level we have at our disposal conditional and unconditional frequencies concerning several factors – e.g. smoking, exposure to other carcinogenic substances, exercising, stress – on the basis of which probabilities in the single case are allocated for the purpose of diagnosis or causal attribution. At both levels causal claims are probabilistic, but how is probability to be interpreted?

3. Interpretations of probability

In the philosophy of probability several interpretations have been advanced and crucial objections raised. Available interpretations are the classical, logical, frequentist, propensity, subjective Bayesian, empirically-based Bayesian, and objective Bayesian. In the following I shall not present and discuss all of them. Excellent and exhaustive introductions to philosophical theories of probability are those of Gillies (2000), Hacking (1975) and Hajek (2003).

Instead, I shall focus on Bayesian approaches only. Firstly, I'll present the basic assumptions behind Bayesianism and then illustrate in the differences between the subjective, empirically-based and objective Bayesian interpretations of probability. This overview is meant to give the reader enough background to understand the choice of an empirically-based or objective Bayesian interpretation for quantitative causal analysis.

There are two main assumptions behind Bayesianism. First, scientific reasoning is essentially reasoning in accordance with the formal principles of probability theory and second, Bayesianism allows learning from experience. The formal apparatus of probability theory, axiomatised by Kolmogorov in 1933, serves to impose coherence constraints on rational degrees of belief and to use conditionalisation as a fundamental probabilistic inference rule for updating probability values according to Bayes' theorem.[1] Thus, Bayesianism allows inductive reasoning from data, namely examining probabilities of hypotheses in the light of data.

Bayesianism, as an epistemological position about scientific reasoning, is accompanied by an interpretation that takes probabilities to be rational degrees of

---

[1] The axioms state that (i) probabilities are non-negative real numbers, (ii) every tautology is assigned value 1, and (iii) the sum of the probabilities of two mutually inconsistent sentences is equal to the probability of their disjunction. The conditional probability of A given B is written P(A|B) and is defined as $P(A \mid B) = \dfrac{P(A \& B)}{P(B)}$, with P(B)≠0. Bayes' theorem follows from the axioms and from the definition of conditional probability. It governs the inversion of a conditional probability and relates the posterior probability of B given A to the probability of A given B, provided that the prior probability of A and B are known or that a conventional procedure to determine them is accepted. Formally, Bayes' theorem states that: $P(B \mid A) = \dfrac{P(A \mid B) \cdot P(B)}{P(A)}$, if P(A)>0. See Howson and Urbach (1993) for a lucid exposition.

belief. In this interpretation, also known as *subjective interpretation* or *subjective Bayesian interpretation*, probabilities are quantitative expressions of an agent's opinion, or epistemic attitudes or anything equivalent. First advances are due to de Finetti (1937) and Ramsey (1926). In this approach probabilities are typically analysed in terms of betting behaviour, namely probabilities are identified with the announcement the betting odds that a rational agent is willing to accept. A *Dutch book* (against an agent) is a series of bets, each equally acceptable to the agent but which collectively guarantee her loss, whatever happens. Two Dutch book theorems then follow. First, if an agent's subjective probabilities violate the probability calculus, then she is liable to a Dutch book, and second, if an agent's probabilities conform to the probability calculus, then no Dutch book can be made against her. An agent is then called rational when no Dutch book can be performed against her. That is to say, obedience to the probability calculus is a necessary and sufficient condition for rationality. It is typically objected that a subjectivist account leads to arbitrariness. In fact, two agents may assign different probability values to the same event and be equally rational, provided that they do not violate the axioms of probability. A solution to the objection of arbitrariness is attempted in the empirically-based and in the objective Bayesian interpretation, which I introduce next.

In a nutshell, these two interpretations require that further constraints be satisfied before an agent's degrees of belief can be deemed rational. Early proponents are Salmon (1967) and Jaynes (1957). Two types of constraints are to be distinguished: empirical and logical. Information and lack of information, respectively, ought to be taken into account in shaping degrees of belief. On the one hand, Salmon emphasises the role of empirical constraints and requires knowledge of relative frequencies to assign prior probability values.[2] On the other hand, Jaynes goes beyond this empirically-based approach and puts forward a maximum entropy principle, which might be thought of as an extension of the principle of indifference.[3]

Thus, whilst the empirically-based interpretation contents itself with the adoption of empirical constraints, namely knowledge of observed frequencies is sufficient to shape degrees of belief, the objective Bayesian interpretation requires that *both* empirical and logical constraints be satisfied. For a novel development of the objective Bayesian approach see Williamson (2005).[4] Also, although both the empirically-based and objective Bayesian interpretations shape degrees on belief using knowledge of observed frequencies, the two significantly differ in

[2] The frequency interpretation is traditionally placed among the "objective" interpretations. Objective interpretations, unlike subjective ones, take probabilities to be quantitative expressions of some features of the world, not of our knowledge or belief about them. A simple form of the frequency interpretation states that the probability of an attribute A in a finite reference class B is the relative frequency of the actual occurrence of A within B. Further developments of the frequency interpretation are due to von Mises (1928) and Reichenbach (1935), who consider infinite reference classes and identify probabilities with the limiting relative frequency of events or attributes therein.

[3] The principle of indifference states that whenever there is no evidence favouring one possibility over another, these possibilities have the same probability.

[4] It is worth noting the shift of meaning of "objective" in Williamson's account. Traditionally "objective" is synonymous of "physical" and interpretations such as the frequentist one are labelled "objective" exactly because they refer to some physical properties of the world. In Williamson's account, instead, the term rather means "non-arbitrary".

that the objective Bayesian approach requires choosing the middling or most equivocal probability value in case of lack of evidence, i.e. of knowledge of observed frequencies (see Williamson 2006).

It is worth noting that Bayesian interpretations, whether subjective, empirically-based or objective, interpret single-case rather than generic probabilities. In fact, degrees of belief are oft associated with bets and a bet in a generic outcome, such as a relative frequency, does not make sense.

4. Twofold causality, twofold probability

As I mentioned at the end of section 2, causal inference falls into two category. In the first category (*generic*), we are interested in establishing causal relations that hold for the population. In the second category (*single-case*), we focus on a particular individual, as is the case in diagnosis or causal attribution.

In the first type of inference, population-level data is collected and analyzed by means of statistical models. Different statistical models, e.g. structural models, are designed to infer relations between variables from large data sets, and these relations will be deemed causal, roughly, if they are sufficiently stable. For instance, generic causal inferences aim at establishing whether tobacco consumption is causally related to cancer, whether marriage dissolution is affected by migration behaviour or the other way round, whether neighbourhood environment influences individuals' health or wealth, etc.

In the second type of inference, the key question is how to combine causal knowledge gathered from population-level data with specific knowledge about a particular individual. For instance, in the case of diagnosis the problem is how to combine generic causal knowledge of diseases with an individual's symptoms, DNA profile and medical history to come up with a diagnosis particularised to that individual. As for causal attribution, again we have to combine generic causal knowledge and knowledge about a particular individual; for instance, to establish whether tobacco consumption or exposure to asbestos caused cancer in a particular patient, we need to correctly apply population risks and combine them with the personal and medical history of the patient.

It is worth noting that although both categories of causal inference are essentially probabilistic they state different things. A generic causal claim posits an average causal relation which is supposedly valid for the majority of individuals exposed to the factor within the population. As for single-case causal claims, two meanings ought to be distinguished: the first meaning is predictive – e.g. your smoking now makes you more likely to develop cancer in the future – and the second meaning is retrospective – e.g. it is likely that your smoking in the past caused you to develop cancer. Thus, single-case causal claims do not state frequency of occurrence but express a belief, in particular a rational degrees of belief, about what did or will happen. This bipartite distinction of causal claims, which for brevity I call *twofold causality*, suggests adopting accordingly a twofold conception of probability. In other words, because generic causal claims state frequencies of occurrence, they apparently need a frequency interpretation of probability. On the other hand, because single-case causal claims state credence about future effects of causes or about past causes of effects, probabilities apparently need an interpretation in terms of rational degrees of belief. Moreover,

because single-case causal statements are informed by population-level causal knowledge, degrees of belief in the single case seem to be *empirically based* upon frequencies stated in the generic causal claim. To account for these intuitions, we need an interpretation able to combine frequencies *and* degrees of belief. The soundness of such combination depends, in the first place, on the plausibility of a *dual* concept of probability, which I shall discuss next.

5. Probability as a dual concept

Probability is a dual concept having an objective and subjective side. This is not tantamount to saying that a principle of tolerance should equally allow all interpretations. Rather, I will defend the idea that the two sides of the concepts serve different purposes.

The duality of the probability concept is analyzed at length in Hacking's *The Emergence of Probability*. Hacking maintains that ever since the first formulations of the probability theory, at the time of the famous epistolary exchange between Pascal and Fermat, "probability" was already meant as degree of belief *and* as the tendency of a chance device to display stable relative frequencies. Hacking's historical thesis is challenged, however. According to Gillies (2000 : 18) the emergence of the duality appeared some time later. It surely started with Laplace, and according to Daston (quoted in Gillies 2000 : 10) the duality traces back to Poisson, Cournot, Ellis. Whatever the correct tenet – Hacking's or Gillies' historical claim – it is a matter of fact that the duality of the concept of probability has a long history. As we shall see later, a number of recent accounts employ *both* the subjective and objective concept of probability too.

The "Janus-faced" aspect of probability, as Gillies calls it, thus appears to be tenable. That is to say, a pluralist view is defensible and different interpretations may better fit different contexts. Gillies himself is a defender of pluralism. He reinforces the pluralist view of probability by arguing that there are two broad areas of intellectual study which require different interpretations of probability. A *subjective* notion of probability is appropriate for the social sciences, whereas an *objective* notion is appropriate for the natural sciences (2000 : 187 ff). Notice, however, that the kind of pluralism here advocated differs from Gillies' in that it requires two distinct concepts of probability in the same domain and hinges on the generic—single-case distinction for causal claims rather than on the domain of application. In other words, this form of pluralism, unlike Gillies', would allow subjective probabilities in the natural sciences (be the causal claim at stake single-case) and for objective probabilities in the social sciences (be the claim at stake generic).

To show how "Janus-faced" probability is involved in contemporary approaches, I will borrow Salmon's (1988) distinction between *frequency-driven* (F-D) accounts of subjective probability and *credence-driven* (C-D) accounts of objective probability. Let me spell out the meaning of F-D and C-D accounts first, I will then go through some logical and subjectivist approaches in order to highlight that they share two characteristics: (i) they employ a twofold concept of probability, and (ii) to shape subjective probabilities they resort to frequencies.

The difference between F-D approaches and C-D approaches can be stated as follows. Given that there are two kinds of probabilities – says Salmon (1988 : 15

ff) – the question to be addressed is how they relate to each other, that is, we have to understand what is the relationship between subjective probabilities and frequencies. In F-D accounts, frequencies play a major role in determining subjective probabilities, whereas C-D account rather go the other way round: objective probability is based on belief. To borrow Salmon's words, the whole point is to make clear whether objective or subjective probability is in the driver's seat, so to speak.

Salmon discusses Ramsey's approach as paradigmatic of the F-D account, and Mellor's and Lewis' as paradigmatic of the C-D account. Salmon obviously sympathises with F-D accounts, as indeed I do. Yet, we are not the only two partisans. I'll now review Carnap's logical account and Salmon's and van Fraassen's Bayesian approaches to show that they ultimately rely on frequencies in order to shape logical and subjective probabilities respectively. I will also discuss Lewis' Principal Principle as it is instructive to emphasise the non arbitrariness of an empirically-based or objective Bayesian approach.

Carnap (1950). In *The Logical Foundations of Probability* Carnap distinguishes two concepts of probability (§9, §10A and §10B). *Probability₁* denotes the weight of evidence or the degree of confirmation, and *probability₂* denotes relative frequency. Probability$_1$ can be explicated in three ways: (i) as a measure of evidential support given to a proposition h in the light of a different proposition e (§41A); (ii) as a fair betting quotient (§41B); and (iii) as an estimate of relative frequency (§41C and §41D). Carnap's strategy is to explicate the concept of probability$_1$ as estimate of relative frequency (iii) *via* the concept of betting quotient (ii), in turn explicated *via* the measure of evidential support (i). For instance, suppose that the relative frequency of the attribute M in a class to which b belongs to is known to be r, then the fair betting quotient for the hypothesis that b is M and hence the probability$_1$ of this hypothesis is r.

In the opening of §8 Carnap makes clear that the concepts of confirmation he will deal with are semantical and logical. In particular, the quantitative concept of confirmation, i.e. *probability₁*, has two arguments – the hypothesis and the evidence – although the latter is oft omitted and in §10A Carnap goes on in stressing the importance of the evidential component of probability$_1$.

Consider now Carnap's *c*-function. The result of an inductive inference has the structure of the *c*-function: $c(h|e) = q$, where propositions h and e express respectively the hypothesis and the evidence, and q is a real number in the interval [0, 1] obeying the axioms of probability theory, i.e. q is a probability value; q here represents the *degree of confirmation* in the hypothesis h on the basis of the *evidence e* (see §8 and §55). It is important to bear in mind that the evidence e represents available experimental or observational evidence, viz. what we *know* about the world and on the basis of which probability values in inductive inferences are shaped upon. Now, if e is also expressed in probabilistic terms, then it has to be interpreted as probability$_2$, that it is to say, probability$_1$ is shaped on frequencies in e. Carnap reiterates this ideas in §§ 41-42 and in §42B says:

"Many writers since the classical period have said of certain probability statements that they are 'based on frequencies' or 'derived from frequencies'. Nevertheless, these statements often, and practically always if made before the time of Venn, speak of probability$_1$, not of probability$_2$. In our terminology they are probability$_1$ statements referring to an evidence involving frequencies. […] in

these cases the probability is determined with the help of a given frequency and its value is either equal or close to that of the frequency".

Salmon (1967). In *The Foundations of Scientific Inference* Salmon tackles the old problem of induction. Induction was famously criticised in the *Enquiry* by Hume, who was seeking an answer to the question of how we acquire *knowledge* of the unobserved. Salmon then draws a distinction between knowledge and belief: knowledge, unlike mere belief, is founded upon evidence, that we need to provide a rational justification for it. Salmon is here raising a *logical* problem: the problem of understanding the *logical* relationship between evidence and conclusion in correct inductive inferences. As is well known, inductive inferences cannot establish true conclusions but only *probable* conclusions from true premises. Because we are dealing with *probable* statements, we also have to provide an interpretation for them. Salmon considers two basic meanings of the concept of probability (1967 : 48-50): as frequency and as rational degree of belief.

The frequentist concept, he argues, is not a suitable interpretation of probabilistic results in inductive inferences. The reason is this. Given that under the frequentist interpretation "the probable is that which happens often and the improbable is that which happens seldom" (1967 : 48), if we claim that inductive inferences are probable in this sense, we would be claiming that inductive inferences with true premises often have true conclusions, although not always. However, Hume's critique of induction has proved two things: (i) inductive inferences cannot establish their conclusion as true, and (ii) inductive inferences cannot establish conclusions as probable in the frequentist sense. Instead, the concept of probability as degree of belief is more promising. In particular, *rational* degrees of belief work better in making clear the meaning of "probable" in conclusions of inductive inferences: to say that a statement is probable means that one would be rationally justified to believe it and that the statement is supported by evidence. Moreover, in inductive inferences, rational degrees of belief are *objectively determined* by evidence. Evidence supports a statement depending on the inductive rules we adopt; according to Salmon, induction by enumeration is the basic inductive rule for this purpose, and allows us to infer the limit of the probability value from the virtually infinite sequence of possible outcomes, i.e. the limit of the relative frequency (1967 : 96-98). The problem now is how we use this evidence in inductive inference to confirm hypotheses.

The solution comes from the probability calculus itself: Bayes' theorem. Bayes' theorem, says Salmon (1967 : 117), "provides the appropriate logical schema to characterise inferences designed to establish scientific hypotheses". However, Bayes' theorem poses difficulties of interpretations. The formal scheme of Bayes' theorem requires prior probabilities – but what are those probabilities? Salmon's answer is: frequencies. "[…] The frequency interpretation of probability can be used to approach the prior probabilities of scientific hypotheses". In sum, rational degrees of belief are based upon the frequencies.

Van Fraassen (1983). In his article on the justification of subjective probabilities[5], van Fraassen distinguishes two uses of the concept of probability. The first refers to the frequency interpretation: probability statements are about

---

[5] In this paper van Fraassen uses a slight different terminology. He oft uses "personalistic" for "subjectivist" and "attitudes" or "opinions" for "rational degrees of belief".

actual frequencies of occurrence of events. The second refers to the subjective interpretation and serves to formulate and express our opinion and the extent of our ignorance concerning matters of fact. According to van Fraassen, any satisfactory account of probability should explicate both uses as well. However, whilst proponents of Bayesian views have done quite well, frequentists haven't. On the one hand, Bayesians have successfully shown that obedience to the probability calculus provides (at least) necessary criteria of rationality. On the other, frequentists have never succeeded in meeting major critiques of their failure to account for the subjective use of probability. As van Fraassen puts it (1983 : 295), he will attempt to redress the balance and to do so he will try to demonstrate that the observance of the probability calculus in expressions of opinion – i.e. rational degrees of belief – is equivalent to the satisfaction of a basic *frequentist* criterion of rationality: potential calibration.

According to van Fraassen rational degrees of belief, once expressed, are evaluated in two ways. One question is whether they are *reasonable* and the another is whether they are *vindicated*. From a Bayesian standpoint – I shall recall – subjective probabilities are equated with the announcement of the betting odds the agent is willing to accept. Thus vindication consists in gaining, or at least not losing as a consequences of such bets. Dutch book theorems then state that vindication will be a priori precluded if and only if probability values do not satisfy the probability calculus. This way coherence is a minimal criterion of reasonableness connected with vindication, in particular, the possibility of vindication is taken as a requirement of reasonableness. However, we can adopt a different strategy to evaluate reasonabless and vindication of degrees of belief: explicate vindication in terms of calibration, and the possibility of vindication in terms of potential calibration.

Simply put, calibration describes the behaviour of a forecaster. A good forecaster should be informative, i.e. probability values assigned to a statement should approach 0 or 1, and well calibrated, where calibration is a measure of how reliable the forecaster is: the higher the frequency of true predictions, the more reliable the forecaster. A forecaster, then, will be perfectly calibrated when she chooses the correct reference class and estimate frequencies that happen to be correct. Potential calibration, or frequency coherence, concerns the extension of the set of propositions beyond the initial set to which we attached probability values and that turned out to be calibrated. The idea of being potentially calibrated is that it is possible to vindicate rational degrees of belief *a priori*. This is possible, according to van Fraassen, because obedience to the probability calculus is equivalent to this frequency-coherence criterion. Therefore, a forecaster can safely shape rational degrees of belief on frequencies insofar she stays potentially calibrated. In other words, van Fraassen is claiming that degrees of belief can be given a frequentist vindication, and in this way he binds subjective probabilities to the frequencies.

Lewis' Principal Principle (1971). Lewis' *Principal Principle* is not a F-D account but rather a D-C. Thus the Principle (Lewis 1971 : 266):

> Let *C* be any reasonable initial credence function. Let *t* be any time.
> Let *x* be any real number in the unit interval. Let *X* be the proposition
> that the chance, at time *t*, of *A*'s holding equals *x*. Let *E* be any

proposition compatible with X that is admissible at time t. Then, $C(A|XE) = x$.

For instance, if A is the proposition that a coin tossed at time t will fall heads, X is the proposition that the chance of A at time t is x, and E is the available evidence that does not contradict X, the Principal Principle then says that x equals the actual degree of belief that the coin falls heads, conditionally on the proposition that its chance of falling heads is x. In other words, the chance of A equals the degree to which an agent believes in A.

The difference between F-D accounts and the Principal Principle is subtle but fundamental. On the one hand, according to F-D accounts I surveyed, knowledge of objective probabilities is used to determine a reasonable degree of belief; that is to say, subjective probabilities must ultimately be based upon objective probabilities, viz. upon knowledge of observed frequencies. On the other, in the Principal Principle the chance of an event is the *chance* of the truth of the proposition that holds at just those worlds where *this* event occurs. It is apparent that Lewis' concept of objective chance widely differs from Popperian propensities[6] in that it is time-dependent and, mostly, world-dependent (remember that Lewis is the proponent of possible-worlds semantics for counterfactuals). Nevertheless, Lewis' objective chance can be assimilated to propensities in the intent of expressing intrinsic characteristics of the world.

Chance, in Lewis' proposal is not used to *shape* subjective probabilities – it is entirely and uniquely determined once our *credence* in the truth of the corresponding proposition is fixed, and evidence does not contradict this credence. Simply put, if our credence in A is x, the chance of A is simply x.

6. The case for frequency-driven accounts

The way in which the Principal Principle tie the objective and the subjective side of Janus-probability boldly differs from the strategies of F-D approaches. Indeed, it is this very same difference that vindicates the choice of an empirically-based or objective Bayesianism. As Salmon points out (1988 : 21), according to the Principal Principle, it seems possible that different agents, with different initial credence functions, will assign different *objective* chances to the truth of A. Agreed, different agents may have different estimates of the chance of A, or different degrees of belief about the chance of A, but the Principal Principle, since it allows different assignments of *objective* probabilities, opens the door to arbitrariness. Frequency-driven accounts - whether empirically-based or objective Bayesian - do not run this risk.

---

[6] Popperian propensities (see Popper 1959) are tendencies or dispositions to produce a particular result or outcome on a specific occasion (for instance, an experiment). Gillies (2000 : 125-126) characterises the propensity theory as any objective but not frequency theory having the following traits: (i) probability is introduced as a primitive undefined term characterised by a set of axioms (Kolmogorov's axioms) and (ii) probability is connected with observation in some manner more indirect than the definition in terms of frequencies. Gillies then distinguishes between a *long-run* propensity theory and a *single-case* propensity theory. In the former probabilities are associated with repeatable conditions and produce, in the long run, frequencies that are approximately equal to the propensities. In the latter propensities produce particular outcomes in specific set ups. According to Gillies Popper was holding both versions.

Let me refer once more the Principal Principle to underline the substantial difference of the usage of objective and subjective probabilities in twofold causality. In the single-case the aim is not to claim credence about chance, e.g. credence about the probability of the causal factor "smoking" in producing lung cancer in Harry. Rather, it is a rational degree of belief in the *hypothesis* that Harry's smoking caused him to develop cancer, given the available evidence about the generic causal claim. In other words, the support of the hypothesis in the single case is based on knowledge about frequencies that hold at the generic level. In sum, such rational degrees of belief are *frequency-driven*.

However, a couple of objections might be raised. The first comes from the staunch objectivist and the persuaded realist. Despite the differences between Bayesianisms, and despite I eventually favour a *frequency-driven* version, it is a matter of fact that Bayesian interpretation of probability deal with rational degrees of belief. On the other hand, rational degrees of belief are features of an agent's mental state, i.e. they are in sharp contrapositions with objective probabilities. So, here is the threat: does the adoption of a subjectivist perspective lead to drop the ambition to acquire knowledge about the world, notably, about causal relations? My reply, in a nutshell, is: no, as long as subjective probabilities are shaped on the basis of available evidence.

Let me argue more widely. To understand why Bayesian approaches do not necessarily lead to an antirealist position concerning causal relations let us have a look again at the Carnapian c-function. As mentioned above, q is the degree to which a piece of evidence e supports or confirms a given hypothesis h. The concept of degree of confirmation ca be lodged within the subjectivist framework because it is more akin to a feature of an agent's mental state rather than an objective feature of the world. Now, I would like to raise the hint: is everything subjective in the c-function? What does exactly the evidence e state? Evidence e represents the experimental or observational evidence, in other words, what we *know* about the world, and what we shape subjective probabilities upon. Differently put, subjective probabilities are not devoid of empirical content as long as they are dependent on empirical constraints.

The second objection, instead, is raised by the staunch subjectivist. Although she would rejoice at the adoption of rational degrees of belief, she would also argue that there is a subjectivist approach that does not employ frequencies at all: this is the approach proposed by de Finetti (1937) and Savage (1954), better known as rational decision-making. Thus, she might wonder why other subjectivists – namely, objective or empirically-based Bayesians – need to introduce frequencies in their accounts. Isn't it just a pedagogical need?

Let us consider, for instance, economic contexts. The force of rational decision-making approach seems to lie in the fact that only one concept of probability – the subjective one – is employed, and that subjective probabilities are sufficient for one's decision. Without going through the technicalities of rational decision making, the very basic idea turns around the concept of *utility*[7], and the general rule for decisions prescribes to maximise the expected utility. That

---

[7] In economic contexts, "utility" measures the degree of satisfaction gained from consuming commodities, i.e. goods and services. Once this measure is specified, one can talk of increasing or decreasing utility, and explain economic behaviour in terms of the attempt to maximise one's utility.

is, this rule says to choose actions for which the estimate of the resulted utility has its maximum. Rational decision-making seems to be applicable even in cases in which the agent does not know beforehand the values of probability – i.e. the relative frequencies – for some events, precisely because she can make *subjective* assignments. That is to say, priors[8] can be assigned arbitrarily (as long as coherence is preserved) and we can, therefore, get rid of frequencies. Thus Savage's (1954) and de Finetti's stance.

Indeed, Carnap (1950 : § 50-51) is aware of the problem, and his solution seems quite accurate. In rational decision-making we can do without frequencies, he argues, *if* inductive logic is accepted. This is for two reasons. The *first* concerns the interpretation of probability$_1$. Carnap can accept that practical decisions be made upon sole knowledge of probability$_1$, because probability$_1$ might be interpreted as estimate of probability$_2$ (§ 41D). And it is true that if the agent actually knows the value of probability$_2$, then the corresponding value of probability$_1$, with respect to this evidence, simply equals the value of probability$_2$, namely the known relative frequency (§ 41C). Nevertheless, the problem still remains, because values of probability$_2$ are unknown in the great majority of cases concerning ordinary economic decision.

And here comes the *second* reason. Values of probability$_1$ are not said to be unknown in the same sense in which probability$_2$ are. That a certain probability$_2$ value is unknown means that we do not have sufficient *factual* information for its calculation. Whereas a probability$_1$ value cannot be unknown in the same sense; according to Carnap, a probability$_1$ value is unknown in the sense that a certain logico-mathematical procedure has not been accomplished yet. The most common situation is that the only available information concerns the frequency of a property *M* pertaining to an observed sample, hence the relative frequency of *M* in the whole population is unknown. Nonetheless, what is still possible to do is to calculate the probability$_1$ of a hypothesis which ascribes *M* to an unobserved individual. This probability$_1$ value is the *estimate* of the unknown probability$_2$ at stake. And this probability$_1$ value will be enough as a basis for the agent's decision. So to go back to the subjectivist's question, do we really need frequencies? It seems that, even though subjective probabilities may be a sufficient basis for decision, in the ultimate analysis those subjective values are formed upon frequencies, *pace* de Finetti and Savage. So far, so good.

Nonetheless, the staunch subjectivist can still play her last card: *exchangeability*. De Finetti's exchangeability argument (1937) claims that *all* probabilities are subjective, and that even apparently objective probabilities can be explicated in terms of degrees of belief. Briefly and very informally, a sequence of random variables $X_1, ..., X_n$ is exchangeable if, for any fixed *n*, the new sequence $X_{i1}, X_{i2}, ..., X_{in}$ has the same joint distribution no matter how *i1,...,in* are chosen.

Let Y_n be the average of any *n* of the random variable X_i, namely:

$$Y_n = \frac{X_{i1} + X_{i2} + ... + X_{in}}{n},$$

---

[8] Priors are probability values assigned to an event of hypothesis in the absence of evidence or before evidence is collected. Posterior probabilities are then probability values computed by means of Bayes' theorem taking into account evidence.

since sequences are exchangeable, it does not matter which sequence *i1, i2, ...,in* is chosen. De Finetti then shows that the different distributions thus generated will tend to a limit as *n* tends to infinity. That is to say, different agents may start with different prior probabilities, but, as evidence is accumulated, their posterior probabilities will tend to converge, thus giving the *illusory* impression that objective probability exists; therefore, de Finetti interprets his mathematical result as showing that we can get rid of objective probability. In his view, objective probability is *metaphysical* in character, and the exchangeability argument could be accepted if one wanted to raise doubts against metaphysical propensies.

Exchangeability might be challenged, however. For instance, Gillies (2000 : 77 ff) takes de Finetti's argument as a reduction from the objective notions of (objective) probability and independence[9], to the subjective notions of (subjective) probability and exchangeability. He then argues against this reduction by saying that the concept of exchangeability is actually parasitic on that of objective independence and, consequently, redundant. Let me explain this point more thoroughly.

As we have seen earlier, according to the subjective interpretation, different agents, although perfectly reasonable and having the same evidence, may have different degrees of belief in the same event or hypothesis and the mathematical theory of probability provides a tool to measure those degrees of belief. However, next to subjective probabilities there seem to be *objective* probabilities too; for instance, the probability that an unbiased die will show an even side or the probability that a particular isotope of uranium will decay in a year seem to be objective rather than subjective. De Finetti's exchangeability argument is meant to show that even these probabilities are *apparently* objectives, apparently because even if agents adopt a strict subjectivist view, as long as they update prior probability values by Bayesian conditionalisation, they will come to agree on their posterior. Gillies' critique of exchangeability (2000 : 69 ff) has two lines of argument. The first one questions the use of Bayesian conditionalisation and can be summarised as follows.

Prior probability functions will in all cases be based on general assumptions about the nature of the situation under study. If these assumptions are correct, then updating prior probabilities by Bayesian conditionalisation will yield reasonable results. However, if these assumptions are wrong, the prior probability values as well as the posterior will be inappropriate. In this case we won't just update the priors but we'll need to choose drastically new prior probabilities values for the event or hypothesis. Therefore, allowing changes in probability values only by Bayesian conditionalisation as de Finetti does is just too restrictive.

The second line of argument questions the general assumptions on which prior probabilities values are based. Gillies (2000 : 75-77) claims that the subjective concept of exchangeability is equivalent to the objective concept of independence. In a strict subjectivist framework as de Finetti's, we can get rid of independence. However, says Gillies, from an objectivist viewpoint we can apply exchangeability *only* when we are in a case of independence. The trouble is that there are many situations in which the outcome of an event is strongly dependent on the outcome of a previous event and here exchangeability will deliver

---

[9] If two events E and F are independent then the joint probability P(E&F) is equal to their product, i.e. P(E)xP(F).

erroneous results. Gillies concludes, we are not able to reduce independence to exchangeability but rather the other way round. In other words, to use exchangeability safely, we have to know first that we are in a situation of independence.

It seems to me that exchangeability is not, after all, a decisive argument against the objective or empirically-based Bayesian interpretation, and particularly against an epistemic use of frequencies. Empirically-based and objective Bayesianism do not "reify" objective probabilities into metaphysical propensities. Objective probabilities have a preferred frequentist interpretation, and there is nothing "metaphysical" in this. Relative frequencies are instead known by *experience*, and this is what guarantees that even subjective probabilities shaped upon them are not devoid of empirical content. The game is still open.

The subjective Bayesian will then rebut that frequencies are just a pedagogical tool, and that subjective Bayesianism, in ultimate analysis, is not ruled out in principle. Different scientists can allocate priors in different ways and all be equally rational: coherence, according to Dutch book arguments, is also a sufficient condition. I might concede that subjective Bayesianism is not ruled out *in principle*, although in this way subjective Bayesianism makes sense of "learning from experience" – which, recall, is one of the two motivations for the Bayesian framework – only *after* priors have been allocated. Let us ask: did experience teach anything at all before the allocation of priors? It did. And this is why we'd better let our rational degrees of belief be frequency-driven.

7. Conclusion

Quantitative analysis is concerned with making causal inferences. Those inferences fall into two categories. I called the first category *generic* and the second *single-case*. The former establishes causal relations from population-level data that hold at the population-level; the latter focuses on particular individuals and is often informed from the corresponding generic relation. These two types of inference share a common feature: they are both *probabilistic*. Their probabilistic character raises the question of the interpretation of probability, which has been the object of the paper.

The defence of a frequency-driven interpretation hinges upon the distinction of generic *vs*. single-case causal inferences and particularly on the different meanings of the corresponding probabilistic statements. In fact, while generic causal claims state frequency of occurrence, single-case causal claims state credence about what will or did happen. To accommodate this twofold meaning, we need an interpretation that allows using frequencies and degrees of belief. I have then argued that such a possibility is provided by an empirically-based or objective Bayesian interpretation.

However, the following question arises: how do we choose between the empirically-based and the objective version? The answer to this question is lodged into a difference between the two accounts. Under the empirically-based interpretation, I shall recall, the only constraint is constituted by knowledge of frequencies, whereas under the objective Bayesian interpretation the agent must also choose the most equivocal or middling value. Now, this extra-constraint proves to be of particular importance when single-case causal claims are at stake,

for instance in the case of diagnosis. Many social sciences are not directly concerned with the individual, their primary concern being to establish generic causal claims. However, other sciences are concerned with the individual either directly (e.g. medicine) or indirectly (e.g. epidemiology). These sciences aim at establishing generic causal relations that are also meant to inform single-case probability assignment. Agreed, quantitative causal analysis could, in many cases, simply adopt the empirically-based interpretation. However, when it comes to the individual and particularly to taking actions depending on the chosen probability value, the objective Bayesian interpretation definitively performs better as it also gives normative precepts.

References

Carnap R. (1950), *The logical foundations of probability*, University of Chicago Press, Chicago.

De Finetti B. (1937), "Foresight. Its logical laws, its subjective sources, in Kyburg H. E. and Smokler H. E. (1964), *Studies in subjective probability*, Huntington, New York, pp. 194-207.

Durkheim E. (1897), *Le suicide*, Presses Universitaires de France, Paris, 1960.

Gillies D. (2000), *Philosophical theories of probability*, Routledge, London.

Hájek A. (2003), "Interpretations of probability", in The Stanford Encyclopedia of Philosophy (Summer 2003 Edition), Edward N. Zalta (ed).

URL = http://plato.stanford.edu/archives/sum2003/entries/probability-interpret/.

Hacking I. (1975), *The emergence of probability*, Cambridge University Press.

Howson C. and Urbach P. (1993), *Scientific reasoning. The Bayesian approach*, Open Court, Second Edition.

Hume D. (1748), *An enquiry concerning human understanding*, edited by Tom L. Beauchamp, Oxford University Press, Oxford/New York, 1999.

Jaynes E.T. (1957), "*Information theory and statistical mechanics*", in The Physical Review, 106(4), pp. 620-630.

Kolmogorov A. N. (1933), *Foundations of the theory of probability*, Chelsea Publishing Company, 1950.

Lewis D. (1971), "A subjectivist's guide to objective chance", reprinted with postscripts in Lewis D. (1986), *Philosophical papers*, II, Oxford University Press, Oxford, pp. 83-132.

Popper K.R. (1959), "The propensity interpretation of the calculus of probability", *British Journal for the Philosophy of Science*, 10, pp. 25-42.

Quetelet A. (1869), *Physique sociale. Ou Essai sur le developpement des facultés de l'homme*, Muquardt, Bruxelles.

Ramsey F. P. (1926), "Truth and probability", in Kyburg H. E. and Smokler H. E. (1964), *Studies in subjective probability*, Huntington, New York, pp. 23-52.

Reichenbach H. (1935), *The theory of probability: an inquiry into the logical and mathematical foundations of the calculus of probability*, University of California Press (1949), Berkeley and Los Angeles. Trans. Ernest H. Hutten and Maria Reichenbach.

Salmon W.C. (1967), *Foundations of scientific inference*, University of Pittsburgh Press.

- (1988), "Dynamic Rationality", in Fetzer J. (1988) (ed), *Probability and causality*, Reidel Publishing Company, Dordrecht, pp. 3-42.

Savage L.J. (1954), *The foundations of statistics*, John Wiley.

Van Fraassen B.C. (1983), "Calibration: a frequency justification for personal probabilities", in Cohen and Laudan (eds), *Physics, philosophy and psychoanalysis*, Clarendon Press, pp. 295-319.

Von Mises R. (1928), *Probability, statistics and truth*, Allen and Unwin, London, second edition (1957)

Williamson J. (2005), *Bayesian nets and causality. Philosophical and computational foundations*, Clarendon Press, Oxford.

- (2006), "Motivating objective Bayesianism: from empirical constraints to objective probabilities", in Harper W.L. and Wheeler G.R. (eds), *Probability and inference: Essays in honour of Henry E. Kybourg Jr.*, College Publications, London.

Wright S. (1934), "The method of path coefficient", *Annals of Mathematical Statistics*, 5(3), pp. 161-215.