Philosophy of Probability

ALBERTO SARACINO

DOI 10.47473/afrwb0001

INDEX

Philosophy of Probability

1. Introduction

Years ago a colleague of mine, external to the Risk Management Department, asked me, as a risk manager, to explain in plane words the concept of probability. I do not remember exactly what I answered, a sort of standard definition like "the level of possibility of something happening or being true" or "a number that represents how likely it is that a particular thing will happen[1]*. That colleague of mine accepted my answer but was not, of course, particularly pleased with it.*

I have a background in statistics and economics. As a student I thought that the then few courses in history and philosophy of probability, statistics and economics were courses useful to "lighten" the heaviness of those studies, and not worthy of real interest. Now, after a professional career in risk management, a subject heavily based upon applications of the theory of probability, I would give a completely different answer, recalling the words spoken by Bertrand Russel (1872-1970), during a lecture in 1929: "… probability is the most important concept in modern science, especially as nobody has the slightest notion what it means"[2]*.*

Indeed, I changed my opinion on the usefulness of those courses. I now believe that in the areas indicated, as well as in other graduate and undergraduate programs, the study of how some paradigms prevailed over others, of how and why competing approaches and methods were abandoned, would be highly desirable.

In 1933, a few years after Russel spoke those words, Kolmogorov (1903-1987) developed the axiomatic theory of probability, which we continue to use nowadays. While there is a broad agreement on the axiomatic theory, concerning the rules of computation, strong differences of interpretation remain about what probability actually is. In spite of the progress that has been made, a unified paradigm has not yet emerged, and there are still different schools of thought.

Roughly speaking, the philosophical theories of probability can be divided into two broad classes: ***epistemic probabilities****, which concern the beliefs of human beings, and* ***objective probabilities****, which concern features of the material world. Epistemic probabilities have to do with degrees of knowledge and would not exist in the absence of humankind, objective probabilities have to do with the essence of the real world, and would exist even in the absence of humankind.*

[1] Cambridge Dictionary online.
[2] Cited in E. T. Bell (1945: 587), The Development of Mathematics, McGraw-Hill (2nd edition).

In section 2 I briefly recall the origins of probability and the classical theory. In section 3 I describe the different philosophical approaches to probability, according to Gillies' (2000) classification; other authors use different classifications and different wordings[3]. In the final section I deal with a pluralist view of probability that, in my opinion, can be of interest for risk managers in steering risk assessment, risk communication, and risk culture as well.

2. Origins and the Classical Theory

Although ancient Greeks developed an advanced mathematics, their interest was mainly in geometry, less so in algebra and arithmetic. Moreover, they had a poor system for numbers and for arithmetical computations, a limitation that disappeared with the adoption of the Indian/Arabic decimal system. It is difficult to think, for instance, of the formulation of a binomial distribution without the availability of a good algebraic notation.

Only in much more later times has some interest in probability arisen. Although Cardano (1501-1576) and Galileo (1564-1642) wrote pages devoted to mathematical problems concerning dice, we generally identify the 1654 correspondence between Fermat (1601-1665), and Pascal (1623-1662) as the beginning of a mathematics of probability. In those years, and for a long time to come, interest in probability was mainly due to interest in gambling games. Indeed, in their correspondence the two mathematicians dealt with gambling problems, including a problem submitted to Pascal by Antoine Gombaud, known as Chevalier de Méré. Poisson would later say that "... a problem proposed to an austere Jansenist [Pascal] by a man of the world [Gambaud] was the origin of the calculus of probability"[4]. In the 17th and 18th centuries, important developments to the theory of probability came from J. Bernoulli (1654-1705), De Moivre (1667-1754), and Bayes (1702-1761), to mention some of the major authors. Leibniz (1646-1716), although not contributing to the mathematics, did however help lay the groundwork for conceptualizing numerical probability as a degree of certainty, that is, as an epistemic notion with relevance beyond the realm of gambling games, which were guiding and limiting interest in probability.

It was under the impetus of the Enlightenment, with Laplace's (1749-1827) "Philosophical Essay on Probabilities," published in 1814, that the classical theory of probability was outlined, inspired by the principles of universal determinism induced by the recent discoveries of Newtonian mechanics. In the classical approach, probability is computed on the basis and under the condition of a number of equally possible elementary cases. Under this assumption, the probability of an event is

[3] For instance, Mellor (2005: 8) classifies probabilities into three basic kinds: physical probabilities (chances), epistemic probabilities, and subjective probabilities (credences).
[4] Cited in Keynes (1921: v).

derived as the ratio of favorable cases to possible cases. The influence, in classical theory, of the interest in understanding gambling games (dice, coins, cards, roulette wheels, etc.) is evident; moreover, the development of the classical theory of probability, based upon the assumption of equiprobability, was facilitated by the spread of regular dice. In a six-sided regular die the probability of getting any one face is 1/6, but this was not the case using astragalus, a foot bone (usually of sheep or deer), in which only four faces were considered, two with probabilities of about 1/10 each, and two with probabilities of about 4/10 each.

3. Logical, Subjective, Frequency, and Propensity Theories

*The **logical theory** was developed in the early decades of the 1900s, at first in Cambridge and then taken up by the members and associates of the Vienna Circle. The best known of the contributors to the logical theory is certainly Keynes (1883-1946), but prominent figures were also Johnson (1858-1931), Jeffrey (1891-1989), Moore (1873-1958), Russel (1872-1970), and the young Wittengstein (1889-1951). It is no coincidence that these authors were all involved in the group of so-called "Apostles", a kind of exclusive club that brought together the best minds of Cambridge, i.e. the "elite of the elite". The only one to resign from this group was Wittengstein, who later did not even join the Vienna Circle.*

In deductive logic, the conclusion is contained in the premises. If all ravens are black and Bintar is a raven, Bintar is black. This theory extends deductive logic from the deterministic dimension to the probabilistic dimension: "inasmuch as it is always assumed that we can sometimes judge directly that a conclusion follows from a premiss, it is no great extension of this assumption to suppose that we can sometimes recognize that a conclusion partially follows from, or stands in a relation of probability to, a premiss"[5]. In such an approach, influenced by the deductive logic of Russell and Whitehead, probabilities are fixed objectively, although in a Platonic sense, in a world of abstract ideas. Probabilities are degrees of rational belief and not degrees of actual beliefs of individuals that may or may not be rational: probabilities are not subjective.

The assumption of equiprobability in classical theory is taken up by Keynes as an a priori principle, which he calls the <u>Principle of Indifference</u>: "… if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability"[6]. The Principle of Indifference rises a number of paradoxes, and is a weak point of Keynes' logical theory. A simple example is the paradox of water and

[5] Keynes (1921: 52).
[6] Keynes (1921: 42).

wine. Imagine that we mix water and wine and we know that the proportion between the two elements is at most 1 to 3, i.e.:

1/3 ≦ ratio(water/wine) ≦ 3.

Due to the Principle of Indifference, the distribution of the ratio(water/wine) has a uniform probability density function in the interval [1/3,3].

Accordingly, the probability of a ratio(water/wine) less or equal to 2 is:

P[ratio(water/wine) ≦ 2] = (2-1/3) / (3-1/3) = 5/8.

If we compute the probability of the same event using the ratio(wine/water), we have:

P[ratio(wine/water) ≧ ½] = (3-1/2) / (3-1/3) = 15/16.

We obtained two different values of probability for the same event, which is, of course, inconsistent.

The ***subjective theory*** *of probability was developed almost simultaneously and independently by Ramsey (1903-1930) and De Finetti (1906-1985); a later version was developed by Savage (1917-1971) under the wording of theory of personal probability. Ramsey's work dates from 1926, but was not published until 1931, while De Finetti published his first papers in 1930. They never had an opportunity to meet or discuss their theories; Ramsey died very young in 1930, at the age of only 26, leaving majors contributions in the fields of mathematics, philosophy, and economics.*

The subjective theory refutes that there is a single rational degree of belief for all individuals. On the contrary, different individuals, while all rational and possessing the same evidence, can and do have different degrees of belief in the occurrence of the same event. The definition of betting quotients is a key point of the subjective theory. A deeper description of betting quotients is beyond the scope of this essay, it is here enough to say that they have to obey a principle of coherence, i.e. an individual's degrees of belief must comply with a rationality constraint.

If you read about probability, you will often read about "Bayesians" and "Bayesianism". Probabilists that make extensive use of and place emphasis on Bayes' rule are usually called Bayesians. Let's assume we have many individuals with different initial beliefs about the probabilities associated with tosses of regular dice. If they modify their a priori beliefs according to Bayes' rule, after many tosses they will agree on their posterior probabilities. In an objectivist's opinion, such a model of learning supports the idea that there is an objective probability[7]; according to De Finetti, "… all that is happening is that, in the light of evidence, different individuals

[7] The issue of subjective probability versus objective probability is developed in section 4.

are coming to agree on their subjective probabilities"[8]*. Indeed, Bayesianism is not confined within the subjective theory of probability, e.g. there is also what we could call a Logic-Bayesianism, having a major contributor in Jeffrey (1926-2002).*

*The **frequency theory** of probability was developed in the mid-19th century by the Cambridge school of Ellis (1817-1859) and Venn (1834-1923), and may be considered the response British empiricists to the continental rationalism of Laplace and his followers. Subsequently, the frequency theory became popular in the Vienna Circle, especially due to Von Mises (1881-1973) and Reichenbach (1891-1953). While the logical theory of probability is to be regarded as a branch of logic, and in the subjective approach probability is the degree of an individual's belief in the occurrence of an event, for Von Mises probability theory is a mathematical science dealing with " ... problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time"*[9]*. The emphasis here, as opposed to the subjective theory, applicable to single events, is on collectives of events, whose probability exists independently of the individuals who estimate it. Von Mises proposes three classes of such events: games of chance (e.g., a long sequence of tosses of dice), vital data and more generally biological statistics (e.g., mortality tables), and situations occurring in physics (e.g., concerning the molecules of a particular gas). A useful distinction is between an empirical collective, observed in finite numbers in the real world, and a mathematical collective, an unobservable infinite sequence; the issue arises whether an empirical collective can be a legitimate representation of a mathematical collective and, if the case, how many observations we need*[10]*.*

The set of all elementary attributes is called "attribute space" by Von Mises; this is the set we now refer to, in a less precise terminology, as "sample space." The frequency theory according to von Mises is based upon two fundamental empirical laws observed in collectives. In Gillies' (2000: 92 and 95) wording they are:

*<u>Law of Stability of Statistical Frequencies</u> **"**It is essential for the theory of probability that experience has shown that in the game of dice, as in all the other mass phenomena which we have mentioned, the relative frequencies of certain attributes become more and more stable as the number of observations is increased"* [11]*.*

<u>Law of Excluded Gambling Systems</u> "The authors of such [gambling] systems have all, sooner or later, had the sad experience of finding out that no system is able to improve their chances of winning in the long run, i.e., to effect the relative frequencies

[8] Gillies (2000: 70).
[9] Von Mises (1928: 11).
[10] The issue is discussed in Gillies (2000: 90-91).
[11] Von Mises (1928: 12).

with which different colors of numbers appear in a sequence selected from the total sequence of the game" [12].

Having stated these two laws, the frequentist probability of an event is defined as the limit of the ratio of the number of experiments in which it occurred to the total number of experiments performed under the same conditions, as that ratio tends to infinity; the existence of such a limit is assured by the Law of Stability of Statistical Frequencies.

In contrast with the classical, logical, subjective and frequency theories, for which there exists a more or less generally agreed version, the same cannot be said of the ***propensity theory*** *of probability, of which there are various versions by different authors. Originally, this theory was developed by Popper (1902-1994) to address the problem of the probability of single events, a problem that the frequency theory is unable to solve. Popper's solution consists in dropping the concept of admissible sequences of repeated actual experiments (a collective), replacing it with the concept of actual or virtual sequences*, *characterized by a set of generating conditions: "… we have to visualize the conditions as endowed with a tendency or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities, which is precisely what the propensity interpretation asserts"* [13].

Popper himself, over time, modified parts of his first version of propensity theory; other versions are attributable to various authors, e.g. Peirce (1839-1914), Mellor (1938-2020), Gillies (1944-living) and Suárez (1968-living). We can distinguish two major classes of propensity theories: long-run propensity theories and single-case propensity theories. The former are associated with repeatable conditions of an experiment, i.e., the propensity to produce long runs of approximately equal probability; the latter are associated with the propensity to produce a particular outcome in a specific context.

It is noteworthy how, in Popper's elaboration, the propensity theory of probability and the concept of falsifiability can be applied to scientific research. Popper recognizes that with probability a strict criterion of falsifiability is not applicable, but "... a physicist is usually quite well able to decide whether he may for the time being accept some particular probability hypothesis as empirically confirmed, or whether he ought to reject it as practically falsified …" [14]*. The sentence closely fits what statisticians and researchers have been doing since Pearson (1957-1936), Gosset (1876-1937), and Fisher (1890-1962) developed the practices of hypothesis testing. These practices are indeed based upon a principle of falsifiability extended to probability.*

---

[12] Von Mises (1928: 25).
[13] Popper (1959: 35).
[14] Popper (1934: 191).

4. A Pluralist View of Probability and Risk Management

It is understood that the philosophical basis of probability is a complex and multifaceted issue.

*The theories presented in the previous sections can be divided into two broad classes: **epistemic probabilities**, which concern the beliefs of human beings, and **objective probabilities**, which concern features of the material world. Gillies (2010:21) supports the idea that Laplace and the probabilists of the period up to him considered probability as epistemic rather than objective. The logical and subjective theories belong the class of epistemic probabilities, the frequency and propensity theories belong to the class of objective probabilities.*

It is worth noting that the main interpreters of the epistemic approach (Keynes, Ramsey, De Finetti) were primarily interested in applications in the economic sciences, while the main interpreters of the objective approach (Von Mises, Fisher, Neyman, Popper) were interested in applications in the natural sciences (physics, biology).

The authors cited above generally hold a "monistic" view of probability, as they claim that their interpretation of probability is general and applicable in all possible cases. For them Hacking (2010:140) has the wording of frequency dogmatists and belief dogmatists[15]*.*

Next to such dogmatic theories, there are now pluralistic approaches to probability, in which the mathematical calculus has several possible interpretations, each valid in a particular context.

A version of such pluralist views of probability is due to Gillies (2010: 179). Gillies proposes a spectrum from subjective to objective consisting of four stages of probability.

1. ***Subjective probabilities**, which represent the degrees of belief of specific individuals.*
2. ***Intersubjective probabilities**, which represent the degree of belief of a group of individuals that reached a consensus.*
3. ***Artefactual probabilities**, which exist in the material world and are accordingly objective, although as the result of an interaction between human beings and nature.*
4. ***Full objective probabilities**, which exist in the material world independently of human beings.*

[15] Hacking (2001: 131) defines "terrible" the terms "subjective" and "objective" as attributes of probability, because in common thinking objective has a positive connotation ("good"), and subjective has a negative connotation ("bad").

Probabilities in stage 1 and 2 are epistemic, probabilities in stage 3 and 4 are objective. The intermediate stages of intersubjective and artifactual probabilities deserve a closer look.

Intersubjective probabilities are introduced by Gillies (2000: 169-175) as a result of rational discussion, in oppositions to personal probabilities[16]. The bases of intersubjective probabilities lay in the fact that many of our beliefs are social in character, and generally accepted by most of the members of a social group. Crucial conditions are required for a social group to originate an intersubjective probability: members must share a common interest, a flow of information and an exchange of ideas.

The probability of getting heads or tails from the tosses of a coin is artifactual, because it is a probability with regard to a real object (the coin), which, however, is a creation of humans and would not exist in the absence of humankind. The probability of survival for one year of a 40-year-old person is artifactual as well, because it directly relates to a real individual (a human body with its specific biology) of the human species, a probability that would not make sense at all in the absence of humankind. Thus, probability is artifactual both when it is "direct" and relates to an entity (full body, body part, group of people, etc.) of the human species, and when it is "indirect" and relates to a non-human entity that has been created or conditioned by humans.

At the lowest stage of the probability scale, an example of subjective probability is an individual's belief in the success of a certain horse in a race. On the other side, at the highest stage of the probability scale, an example of completely objective probability is the probability of decay per unit time of a particular radioactive atom.

Indeed, assigning a probability to one or another of the four stages is not always smooth and easy. Moreover, there may be difficulties in defining an estimation method even within a category.

Gillies' (2000: 89, 119, 121 and 179) example of the artifactual probability of survival for one year of a 40-year-old person is significant in this regard. For the purpose of estimation, we could use the survival tables of 40-year-old individuals of the same gender, of 40-year-old individuals of the same gender in the state or region of residence, and so on. If the individual smokes two packs of cigarettes a day and plays sports once a week, we could use the tables of 40-year-olds who smoke two packs of cigarettes a day, the tables of 40-year-olds who play sports once a week, or the tables of those who do both. Indeed, we could define as a general principle the principle of

[16] Gillies' definition of intersubjective probabilities is influenced by the concept of paradigm developed in Kuhn T. S. (1962), The Structure of Scientific Revolutions, University of Chicago Press.

considering the class of membership given by the intersection of all possible classes to which the individual belongs, with a numerosity constraint; the availability of data on such a set would still have to be verified. Moreover, even assuming that we can use the "most disaggregated reference class" principle, an uncritical use of this principle may lead us astray. In particular, we might have qualitative information, about the individual in question, worthy of consideration and such as to correct the probability upwards or downwards: diet, longevity of close relatives, more or less wearing profession, and so on.

Probabilities in risk management generally belong to stages 2 and 3 of Gillies' classification, although stage 1 and stage 4 can occur. For a risk manager the differences between epistemic probabilities and objective probabilities do not make a practical difference in ***probability computations***[17]*. Indeed, implementing and running an* ***Enterprise Risk Management*** *(ERM) framework is much more than just computing probabilities; a clear allocation of risk events to epistemic probabilities or objective probabilities can be useful in properly addressing some ERM components, specifically risk assessment, risk communication, and in some respects risk culture as well.*

In ***risk assessment****, a major difference exists between the practices of estimating stage 2 epistemic probabilities or stage 3 subjective probabilities. In the case of stage 3 probabilities, estimation is supposed to concern relatively well-known events, for which some "scientific" evidence is available, as in the example of the probability of survival of a 40-years-old person. Although with the difficulties mentioned in the example, risk assessment should start from the collection of all sources of such available evidence. On the other side, stage 2 probabilities are supposed to lack "scientific" evidence, as in the case of a single or new event, and can rely only on "rational" belief, i.e. logic thinking, rational hypothesis, rational models, and so on. For epistemic probabilities, the effort should be on the research of consensus from all the stakeholders of the risk assessment. It may be the case that professional associations define a standard practice, or regulators release a norm for some risk categories with epistemic probability, and in these situations reaching consensus is straightforward.*

In the case of stage 2 probabilities, ***risk communication*** *should stress the rationality and consensuality of the estimation process, without giving the message of an "objective" probability that does not exists. On the other side, for stage 3 probabilities, risk communication should be focused on the "scientific" reliability of sources, data, and methods.*

---

[17] Peschard et al. (2010: 43): "… the differences between the frequentist and subjective views are good to keep in mind but do not make a practical difference for our [risk] computations".

Finally, ***risk culture*** *should clearly address the difference between epistemic probabilities and objective probabilities. Moreover, risk culture should steer attention towards relevant epistemic probabilities, as in the case, for instance, of VUCA risks*[18]*.*

[18] Volatility, Uncertainty, Complexity, Ambiguity.

Readings

An extensive bibliography is available on the philosophy of probability.

The following introductory texts, although avoiding advanced mathematics, assume some familiarity with the theory of probability:

- *GILLIES D. (2000), Philosophical Theories of Probability, Routledge.*
- *HACKING I. (2001), An Introduction to Probability and Inductive Logic, Cambridge University Press.*
- *MELLOR D.H. (2005), Probability: A Philosophical Introduction, Routledge.*
- *PESCHARD I., BENÉTREAU-DUPIN Y., and WESSELS C. (2023), Philosophy and Science of Risk, Routledge.*
- *SUÁREZ M. (2020); Philosophy of Probability and Statistical Modelling, Cambridge University Press.*

A more advanced and broader picture of the philosophy of probability can be found in:

- *EAGLE A. [editor] (2011), Philosophy of Probability: Contemporary Readings, Routledge.*
- *HÁJEK A. and HITCHCOCK C. [editors] (2016), The Oxford Handbook of Probability and Philosophy, Oxford University Press.*

Classical readings on probability include:

- *DE FINETTI B. (1970) Theory of Probability, Wiley (English edition, 1974).*
- *KEYNES J. M. (1921), A Treatise on Probability, Macmillan (reprint, 1963).*
- *JEFFREYS H. (1939), Theory of Probability, Oxford University Press.*
- *KOLMOGOROV A. N. (1933), Foundations of the Theory of Probability, Chelsea (2nd English edition, 1956).*
- *LAPLACE P. S. (1814), A Philosophical Essay on Probabilities, Dover (English translation of the 6th French edition, 1951).*
- *POPPER K. R. (1934), The Logic of Scientific Discovery, Hutchinson (6th revised reprint of the 1959 English translation, 1972).*
- *POPPER K. R. (1957), The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory, in KÖRNER S. [editor], Observation and Interpretation in The Philosophy of Physics, Proceedings of the 9th Symposium of the Colston Research Society, Bristol.*
- *POPPER K. R. (1959), The Propensity Interpretation of Probability, British Journal for the Philosophy of Science 10: 25-42, reprinted in EAGLE A. [editor] (2011: 488-497).*
- *RAMSEY F. P. (1931), Truth and Probability, in Ramsey F. P., The Foundations of Mathematics and other Logical Essays, Routledge and Keagan Paul, reprinted in EAGLE A. [editor] (2011: 48-71).*
- *VON MISES R. (1928), Probability, Statistics and Truth, Allen and Unwin (2nd revised English edition, 1961).*
- *VON MISES R. (1957), The Definition of Probability, Dover, reprinted in EAGLE A. [editor] (2011: 355-387).*