A Historical and Philosophical Perspective on Probability

Manfred Borovcnik and Ramesh Kapadia

Abstract This chapter presents a twenty first century historical and philosophical perspective on probability, related to the teaching of probability. It is important to remember the historical development as it provides pointers to be taken into account in developing a modern curriculum in teaching probability at all levels. We include some elements relating to continuous as well as discrete distributions. Starting with initial ideas of chance two millennia ago, we move on to the correspondence of Pascal and Fermat, and insurance against risk. Two centuries of debate and discussion led to the key fundamental ideas; the twentieth century saw the climax of the axiomatic approach from Kolmogorov.

Philosophical difficulties have been prevalent in probability since its inception, especially since the idea requires modelling—probability is not an inherent property of an event, but is based on the underlying model chosen. Hence the arguments about the philosophical basis of probability have still not been fully resolved. The three main theories (APT, FQT, and SJT) are described, relating to the symmetric, frequentist, and subjectivist approaches. These philosophical ideas are key to developing teaching content and methodology. Probabilistic concepts are closer to a consistent way of thinking about the world rather than describing the world in a consistent manner, which seems paradoxical, and can only be resolved by a careful analysis.

1 Introduction and Sources

The short history of probability as compared to other mathematical concepts has deep reasons, which differ from those that caused the axiomatic treatment of numbers to occur so much later than Euclid's axiomatization of geometry. One reason behind this tardiness may be the confusion about the purpose of the concepts. Will probability in a mathematical formulation help us to know the future in advance? Or, is probability just a view on the world by which one might get an advantage in

M. Borovcnik (✉) · R. Kapadia
Institut für Statistik, Alpen-Adria Universität Klagenfurt, Universitätsstraße 65, 9020 Klagenfurt, Austria
e-mail: manfred.borovcnik@uni-klu.ac.at

E.J. Chernoff, B. Sriraman (eds.), *Probabilistic Thinking*,
Advances in Mathematics Education, DOI 10.1007/978-94-007-7155-0_2,

acting more consistently? The positive answer to the second question is of limited value to the emotive hope of the first question.

For example, the statement that the probability to roll a six with a die is $1/6$ arises from the application of a mathematical model. To ask whether a die has a physical property to show a six (be it one sixth or an unknown number p) reveals a philosophical bluntness, as if such a property is comparable intrinsically to the weight or the diameter of a coin. In a general philosophical debate, a probability statement as a physical property of the die would be construed to be similar, yet different from the property of the 1 US cent coin to weigh 2.05 gram and have a diameter of 19.05 mm. One major difference is that a probability statement of the die describes a model property, which might be indirectly measured by some ideal procedure but not measured in reality. One way to decide whether to adopt a probabilistic view on a specific die is to investigate the results of tossing it.

Though a sound mathematical foundation was published in 1933, this has not clarified the nature of probability. There are still a number of quite distinctive philosophical approaches which arouse controversy to this day.

There are useful sources for the history of probability like Todhunter (1865), David (1962, early traces), Maistrov (1974, not much before Pascal and Fermat, with a special focus on the Russian school till von Mises), Schneider (1989, till Kolmogorov, in German), Porter (1986, frequentism in the nineteenth century), Daston (1988, classical probability in the enlightenment of the eighteenth century), Hald (1990, 2007), Stigler (1986, till 1900). Philosophical issues are dealt with in Stegmüller (1973), Fine (1973), Barnett (1973), Weatherford (1982), and Hacking (1975, 1990). There are many books on probability itself such as von Plato (1994) and Çinlar (2011). The interested reader should consult these books for more detailed coverage of the ideas underlying probability. We have used these books to source key ideas which are relevant in probability education. Readers should also note that we have used these references rather than the originals, many of which are obscure and in other languages.

For our purpose of teaching probability, we will focus on the following three philosophical theories, which are discussed in more detail later:

- Classical a priori theory (APT); this is developed from the original classical theory.
- Frequentist theory (FQT); this is based on experimental results.
- Subjectivist theory (SJT); this is based on personal belief and linked to Bayes' theorem.

2 From Divination to Combinatorial Multiplicity

2.1 Early Origins in Divination and Religion

Early notions of chance can be found in the ancient cultures of Indians, Babylonians, and Egyptians. David (1962) refers to the astragalus—a bone in the heel of a sheep—as the earliest known object used for games of chance around 3500 B.C. It is possible that primitive dice were constructed by rubbing the round sides of the astragalus until it was approximately flat. Yet, so long as the real bones were used, their natural variation would affect their comparability and thus increase their characteristic to be unpredictable, a key element of chance. The cubes in Babylon of 3000 B.C. made from well-fired pottery were nearly symmetrical dice.

However, there were no systematic investigations—neither of combinatorial multiplicities nor of frequencies to study the results. In her book on the origins and history of probabilistic ideas from ancient times to the Newtonian era, David (1962, p. 21) speculates on why the conceptual progress of probability was so tardy:

> It was fifteen hundred years after the idealization of the solid figure before we have the first stirrings, and four hundred and fifty years after that before there was the final breakaway from the games of chance, if indeed we have really accomplished it even today.

In fact, there were some attempts to judge the likelihood of the outcomes as Cicero in his *De divinatione* (44 B.C.) expresses in the following argument:

> They are entirely fortuitous you say? Come! Come! Do you really mean that? ... When the four dice produce the venus-throw you may talk of accident: but suppose you made a hundred casts and the venus-throw appeared a hundred times; could you call that accidental? (David 1962, p. 24).

In one throw, it is unlikely to have all four sides up with a different result (the venus-throw), and it is even more unlikely to get this result a hundred times in series. The argument was based on the near impossibility of that outcome (not on a combinatorial basis) but in the sense of seeing god's will in such an outcome. Cicero continues to speculate on the character of randomness:

> Is it possible, then, for any man to apprehend in advance occurrences for which no cause or reason can be assigned? (David 1962, p. 24)

There are sound reasons that theoretical arguments were not developed. The Greek philosophy had an ideal of true relations, which prevented the construction of theoretical hypotheses from empirical data. The Greek tradition is illustrated by the "beauty" of Platonic bodies, which have underlying symmetries, and Euclid's geometry, which is based on axioms. Data about variability was ignored as contrary to their ideal of scientific argument. This is an external explanation of the tardy development of concepts for probability. An internal one might be that probability right from the early origins was intimately bound to divination to predict the future (Borovcnik 2011).

Divine judgement at religious ceremonies (at Delphi and elsewhere) was often based on what we would now call games of chance—beans were drawn from a bowl of white and black beans to answer binary questions. Four or five astragali were thrown and the answer to a question was linked to the combinatorial possibilities: for each outcome a structural response to the posed question was prepared. The combinatorial multiplicity may have been implicitly known but links to the outcome were not investigated.

Dice were recognized as a tool to decide fairly or to explore god's will for a decision. From the letters of Lucas we see that—when they could not decide between two persons—they cast lots for them. From the Old Testament there are several places known where a decision was determined by a chance experiment (about 70 places are known to refer to such a use of games of chance, see Logos Bible Software n.d.). For exploring god's will it was not so important that lots were fair.

In the Christian era, starting with the acceptance of Christianity as the only allowed religion (under Theodosius, 380 A.D.), games of chance lost prestige as everything that happens is determined by the will of god. This attitude is expressed by St Augustine (354–430) as we may see from a quotation from David (1962, p. 26):

> [...] nothing happened by chance, everything being minutely controlled by the will of God. If events appear to occur at random, that is because of the ignorance of man and not in the nature of the events.

Such a strict attitude of determination by god is hardly reconcilable with the free will of an individual. Thus Christian philosophy hindered theoretical discussion of randomness.

It is interesting that games of chance did not remain the exclusive preserve of priests; they became an important leisure activity throughout the Roman Empire. The addiction to play such games of chance may nourish a deep desire to explore god's will in everyday situations (the popularity of games of chance is unbroken till today—the business of games of chance is an important branch of applications of probability). At times, Roman emperors forbade games with no success.

Considerable experience would have been gained by priests and other important people from casting dice or drawing beans out of urns for divine judgement at religious ceremonies. In some ways, it is curious that a conceptual breakthrough failed to appear to formalise probability, based on the regularity of the fall of dice. It could be that priests were taught to manipulate the fall of the dice to achieve a desired result, as the interpretation of divine intent, conveyed through their pronouncements. More critically, speculation on such a subject might have brought a charge of impiety in the attempt to penetrate the mysteries of the deity. Moreover, men knew that nature was fickle and while there were elements of regularity such as the rising of the sun every day many other elements of life were hard to predict, not least the weather. Divination was important and a key element of the power of religion and its high priests, as a means of control of ordinary people.

According to Borovcnik et al. (1991, p. 28), one should not overestimate this tardiness of conceptualization as this is also true for other disciplines.

> The development of a causal physical approach to science instead of a deistic one was marked by great controversies, even if today one finds such ideas naturally accepted. Think of Galilei's (1564–1642) troubles with the Pope. Euclidean geometry was not the breakthrough as is generally viewed, neither in axiomatic thinking nor in geometry. It only provided rules for construction but no formal concepts; the status of the parallel axiom was not clarified until the work of Gauss and Lobachevski in the nineteenth century; the concept of continuity of lines and planes was developed even later, though it is implicitly assumed by Euclid. In arithmetic, on the other hand, no sound answer was given to the question of axiomatizing numbers until Peano over 100 years ago. However, a comparable milestone was not reached in probability until 1933.

One further obstacle to conceptual progress might also lie in the following aspect, which focuses on the individual faced with a random situation, be it ruled either by coins or by divine judgement: nothing but the *next* explicit outcome counts. However, this cannot be predicted by any human being. Conceptual progress can be achieved by interpreting this single case as one representative of a series of future (or hypothetical outcomes). Analytical philosophy of science has not come up with any satisfying solution of that problem (as expressed by the comprehensive and extensive research of Stegmüller 1973). But, all the more, for the individual any solution to the series of future outcomes has no attraction as—even if there were more comparable situations to face in future—the individual would always focus on the present situation with all its ingredients including fate, god's decision, impact of outcomes, which will never be the same. In their experiments, Kahneman and Tversky (1979) have shown that even an interchange in gain or loss in a situation is very influential. Thus, the answer to the series is never an answer to the single case. Surprisingly enough, a probability for the series still yields some information for the single case. So, the knowledge about tossing a coin allows a fair decision between two opponents when deciding the sides to play in a football game.

2.2 Emergence of the Rule of Favourable to Possible: Combinatorial Multiplicity

The poem *De vetula* from the thirteenth century may be the first explicit link between combinatorial multiplicity and frequencies of outcomes. For three dice, all 216 possible cases are listed. Of the 16 possible sums, it is advisable to put wagers according to expected gain as "you will learn full well how great a gain or loss any of them is able to be" (Bellhouse 2000, p. 135, cited from Batanero et al. 2005, p. 20); however, this and Peverone's attempt to enumerate the possibilities of a game (Peverone 1558) do not establish the notion of probability. Similarly, Cardano's deliberations in his *Liber de ludo aleae* (approx. 1564) confuse expected value and probability. Cardano discusses how equally easy it is to get any of the numbers in casting one die but then proceeds with an awkward argument about the consequences:

> One-half the total number of faces always represents equality; thus the chances are equal that a given point will turn up in three throws, for the total circuit is completed in six, or again that one of three given points will turn up in one throw. [...] The wagers therefore are laid in accordance with this equality if the die is honest. (Cardano approx. 1564, quoted from David 1962, p. 58)

It is doubtful that Cardano saw a clear relation between empirical frequencies and a theoretical concept like probability.

The main books on the history of mathematics give the credit of the first conceptual approach to probability to a famous correspondence. This suits a romanticised viewpoint, even if the authenticity can be doubted. It is the evil of gambling which seemingly led to the fall of determinism. More generally, people do build up a personal view of chances from daily life and its encounters with chance. Thus it was a century after Cardano that Pascal and Fermat made progress in conceptualizing probability in their exchange of letters in 1654 (published in Fermat 1679; see also David 1962). They discussed and solved two specific problems, de Méré's problem and the Division of Stakes (problem of points). Both problems, which are described below, relate to gambling and fairness, using the idea of proportion, which is the basic building block of probability; yet the use of proportion is not always as straightforward as one might imagine. This is too often neglected when linking these ideas within the educational sphere.

De Méré's Problem Two games are compared. In game 1, the player wins if there is at least one 'six' in four throws of a die; in game 2, the player wins if there is at least one 'double six' in 24 throws of two dice. At first sight, the games seem to be very similar. The solution of Pascal and Fermat is based on an exhaustive enumeration of the sample space (or, fundamental probability set), following the work of Galilei (who took up the three dice of *De vetula*):

$$\mathrm{P}(\text{win in game 1}) = 1 - (5/6)^4 = 671/1296 = 0.518 > 1/2,$$
$$\mathrm{P}(\text{win in game 2}) = 1 - (35/36)^{24} = 0.491 < 1/2.$$

According to Ore (1953, p. 411), de Méré rejected that solution and referred to the equity of proportions: 24 (opportunities to get the desired result, which were seen as favourable cases) to 36 (possible cases) has the same proportion as 4 (opportunities) to 6 (possibilities).

It is romantic folklore that de Méré won a fortune by game 1 and lost everything by game 2. However, a deviation of 0.009 from 0.5 in game 2 requires roughly 13,000 data to detect it (depending on the significance level of a modern statistical test). Second, with the experience of 80 bets per week (roughly 4 nights of gambling with 2 hours on the dice table) a 95% confidence interval for the final balance in game 2 for half a year would be roughly -122 to 53 with an expected value of -34 units. Multiply that with stakes of 100 instead of 1 unit still would result in expected losses of 3,400 and a—probabilistic—upper boundary for the losses of 12,200. A person who can stake 100 per game would not be deprived of all his fortunes if losing 12,000 within half a year.

De Méré's rhetorical rejection of the solution reflects a careless use of the "favourable to possible" rule developed at that time. Indeed, the favourable "cases" refer to repetitions of the experiment and are not a subset of the possible cases. There was a conflict between an enumeration of the sample space and the rule of favourable to possible, which still had to be clarified. Another source of complication might again be a confusion between probability and expected value. While the probabilities are different for both games, the expected number of the desired result is the same in both games—there are 4 trials, each has the expected value of $1/6$, thus the expected number of sixes in game 1 equals $2/3$. In game 2, with 24 trials with expected value of $1/36$ each, the expected number of times a double six occurs, is again $2/3 = 24/36$. We will illustrate the difference of both games by a *scenario* of 100 games (see the table below): in 48 runs of game 1 no six occurred and the bet is lost; for game 2, 51 bets are lost as no double six occurred.

In both series the average number of sixes (double sixes) is 66, which corresponds to the expected value of $2/3$ per run of the game. While the bet is lost in game 2 more often, in game 2 the event "three or more" occurred more often than in game 1. However, there was no extra payment for this excess of double sixes.

How often	6's in game 1	66's in game 2
0	48	51
1	39	35
2	12	11
3 or more	1	3
Average	66	66
Win the bet	52	49

Division of Stakes If Peter and Paul are competing, how should they share the stakes if the series is ended at a time when Peter needs two points and Paul needs three more points to win? Suppose, for example, in a series of 11 Peter has 4 points while Paul has 3 points.

Huygens (1657) incorporated the problem and its probabilistic solution in his *De ratiociniis in ludo aleae*. The perception of the situation with probabilities, once introduced was so convincing that other arguments were defeated. The conclusion is that the stakes should be divided proportionately to the probability of winning in the continuation of games and neither in the proportions 3:2 nor 4:3 (this solution was favoured by Pacioli 1494). Pascal later developed his famous arithmetical triangle as a general method to solve similar problems involving binomial experiments.

Pascal and Fermat's approach sheds light on the correct application of what they termed to be the 'favourable to possible rule', but they made less progress in trying to formally define the concept of probability. They used probability pragmatically as the equal likelihood of outcomes in games of chance, which seemed to be intuitively obvious to them. Hence the emergence of the classical a priori theory (APT) of probability, which later was linked to the principle of indifference discussed below, based on the ideas of Laplace. Games of chance eventually served as a vital link between intuition and developing concepts as well as a tool to structure real phenomena. This view is also supported by Maistrov (1974, p. 48):

> These games did serve as convenient and readily understandable scheme for handy illustration of various probabilistic propositions.

3 Huygens, Bernoulli, and Bayes: The Art of Conjecturing

3.1 Expectation and Probability

A multi-faceted personality in the history of probability is Huygens (born in the Hague in 1629)—the real begetter, the man who synthesized ideas on *probability* in a systematic way. His book *De ratiociniis in aleae ludo* was published in 1657 and was not superseded for over half a century. But strangely his 14 propositions do not contain the word of probability, though he does mention chance. Huygens refers to situations of equal chances and later only to having p chances. And he derives no specific probabilities but proportions of stakes or a price of the situation.

> Proposition 3: to have p chances of obtaining a and q of obtaining b, chances being equal, is worth $\frac{pa+qb}{p+q}$.

In his analysis of the historical development, Shafer (1996, p. 16) explicitly refers to the non-probabilistic embedding of the emergent concepts:

> Pascal, Fermat, and Huygens were concerned with a problem of equity, not a problem of probability. They were pricing gambles, not evaluating evidence or argument. But it did not take people long to draw the analogy.

For Huygens there are two worlds with no direct connection. From games of chance with equally likely outcomes he derives the value of an enterprise (its expectation) as an economic price, a theory of equity. Huygens does not use the word probability to denote the proportion of stakes for a player as his probability of winning. In this approach, Huygens derives recursive rules for expected values if the basic situation is interpreted with probabilities from today's perspective. But he does not use the term expected value (a notion that came into the Latin translation by Francis van Schooten); instead Huygens speaks about the price or true value of a pay-off table.

Probability was also the ingredient of the other "theory" on decisions, on which Huygens worked. Probability was not yet a number but the collection of arguments (pro and con) in order to weigh these arguments. However, since Huygens, the frequency of specific observations becomes a possibility to substitute for other arguments:

> [...] especially when there is a great number of them [...] one can imagine and foresee new phenomena, which ought to follow from the hypotheses [...] and when one finds that therein the fact corresponds to our prevision. (Huygens in Treatise on Light, cited from Shafer 1996, p. 18)

Another application of probability emerged at that time. Huygens established mortality tables and treated frequencies in the same way as probabilities. Moreover, he defined theoretical concepts like the mean life time. These are the first signs of the idea of probability as a concept or theory based on relative frequency (FQT).

However, there was no deeper theoretical argument for such a statement before Bernoulli. Probability plays the role of an undefined elementary concept, which is derived from games of chance. Huygens does not develop any combinatorial methods which are consistent with his view on probabilities which are reduced to equity of cases of the underlying situation; his concept of expectation, recursively applied, yields solutions to various problems.

The first Englishman to calculate empirical probabilities was Graunt (1662). This was a natural extension of the Domesday Book, the most remarkable administrative document of its age, which enabled the King to know the landed wealth of his entire realm, how it was peopled and with what sort of men, what their rights were and how much they were worth. Speculations and research into probability theory did not concern the English, whose interest was in concrete facts.

Graunt's importance both as a statistician and an empirical probabilist lies in his attempts to enumerate as a fundamental probability set the population of London at risk to the several diseases such as are given in the Bills of Mortality recorded by London clerks. This gave further impetus to the collection of vital statistics and life tables. As Graunt epitomizes, the empirical approach of the English to probability was not through the gaming table, but through the more practical and raw material of experience. Graunt continues a tradition already started by Huygens to calculate empirical frequencies and deal with them as if they were probabilities without theoretical justification. For such problems like annuities, a practical solution was needed. A more theoretical problem was to estimate the mortality of diseases. These ideas are the precursors to insurance and risk.

Jakob Bernoulli (1713/1987) culminates this early development when he explicitly uses probability as parts of unity and supports this perception with his extensive work on combinatorics, in which the numerical probabilities of many standard problems (of the time) have been derived. He did not refer to proportions of stakes in games, bets, or economic enterprises, but used probability as a proportion or part of certainty for occurrence. His second achievement is what is now called the law of large numbers; Bernoulli himself called it *theorema aureum*, in which he derives mathematically a relation (a kind of convergence) between equal probabilities and the observed frequencies in the repetition of such games. It is this relation which justifies a more general concept of probability based on observed frequencies. Before this theorem, probability was derived from equal possibilities; after this theorem was proved, the way was open for a new concept of probability, namely probability linked to frequencies, which frees probability from equal likelihood.

Beyond this achievement, Bernoulli extended arguments on how to combine various parts of evidence on the basis of what we would now call independent events. His *Ars conjectandi* (1713)—the art of conjecturing—was published only years after his early death. Though Bernoulli thoroughly explored how to combine the evidence from several events, he did not develop realistic examples to illustrate the conditions that statements have to fulfil in order to fit his theorem. The use of relative frequencies can fail if the conditions of the single experiments are different. This was embedded and hidden in the Bernoulli distribution, which was the basis of his theorem, but was not noticed by the research community who had already gone further.

Bernoulli's philosophical ideas still stick to metaphysical determinism wherein everything is ruled by deterministic laws—be it weather, dice or eclipses of planets. Most phenomena were perceived to be so complex that it would be pointless to study the possible cases. However, his law of large numbers allows the use of observed frequencies for probability:

> [...] to ascertain it from the results observed in numerous similar instances. (Maistrov 1974, p. 69)

3.2 Obstacles and Further Developments

One often learns the most about a concept by exploring situations, in which it fails. Bernoulli had given one possibility to connect probability to relative frequencies but this was only part of his approach. He knew that this could fail either with insufficient data to measure the probability accurately (probability was used as a physical quantity at that time) or because the condition of independence is not satisfied.

Bernoulli's introduction of the new term probability, as being a part of certainty (i.e. a fraction between 0 and 1 instead of proportions of favourable to unfavourable cases) was generally accepted. However, the delay of the publication of his work allowed the community to develop ideas along other lines ignoring his approach that probability could be measured by empirical frequencies with enough data and the relation between frequencies and probabilities resulting from his *theorema aureum*. Thus, de Moivre's *Doctrine of chance* (1718) had a cursory reference to the law of large numbers and only in its second edition of 1738/1967, was the theorem included as a mathematical highlight, yet completely isolated from the rest of the book. Bernoulli's art of conjecturing to combine probabilistic arguments was abandoned, and when Bayes (1763) went back to solve it, his approach was very different. But before discussing Bayes, we describe a key aspect of Huygens' expected value of a situation, which remained an ingredient—the St Petersburg problem.

Cramer corresponded with the Bernoulli family about the following problem, which was published in 1738 as an appendix by Daniel Bernoulli (1738/1954) in the Petersburg Commentaries.

St Petersburg Paradox Two players A and B toss a coin until it shows a 'head' for the first time. If this occurs at the nth trial, then player B pays 2^{n-1} écus to player A. What amount should A pay to B before this game starts in order to make it fair?

If X denotes the amount B pays, then its sample space is (in theory) all natural numbers. The expected gain is $E(X)$, but this is infinite as the series involved diverges:

$$E(X) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + \cdots + 2^{n-1} \cdot \frac{1}{2^n} + \cdots.$$

Thus player A would have to pay an infinite amount of money to B before the game starts.

A modern standpoint would see no problem. If an expected value is infinite this is seen as trivial since a mathematical model of a situation cannot reflect all its properties. The intuitive clash is irrelevant that we cannot play a game for ever. If, in a random experiment, the model used leads to a random variable with no existing expected value, everything is different. The Bernoulli law of "convergence" of the relative frequencies to the underlying probability is no longer valid if the expected value does not exist.

However, we return to the historic situation of the curiosities the problem induced. No one can play such a game as no person has an infinite amount of money.

Yet the expected length of the game is 2 and the associated payoff is 2. We may even be intuitively ready to stake 10 units (historic unit was écu). How can Huygens' expected value deliver a value of infinity in what seems to be a normal situation? At the time, people did not think they were applying a mathematical model to a real situation, or that the model could be inadequate.

Their approach sheds light on the historic perception of probability in the eighteenth century. Probability was not yet anchored by a unified theory, nor was Bernoulli's theorem common-place. Probability was perceived as kind of provability. So the mathematical consequence of getting an infinite value of a game was unacceptable. Much later, Venn formulated a harsh critique. He asserts that no man would pay 50 units as

> neither he nor those who are to pay would be likely to live long enough to obtain throws to remunerate him for one-tenth of his outlay, to say nothing of his trouble or loss of time. (Venn 1888, p. 155)

Venn did not refer to the waiting time until one wins an amount higher than 50 units (an event that has a probability of $1/2^6$). He meant, it would take more than a life to reach a point of time when the net balance of all previous games is positive (or at least zero).

One way to repair the theory is to introduce utility of money to make the expected value finite. A second way out of the dilemma was to introduce moral probability, a magnitude below which any probabilities are ignored. While the first leads to modern approaches of utility it does not solve the paradoxical feature of the Petersburg problem as one could change the payments of the game and still find an infinite expected utility. The time was not ripe to consider the special choice of the utility function merely as one of several possible models. The second approach shifts the dilemma to another level as to when a probability is small enough to be discarded and set to 0. The concept of moral expectation gained much support from Buffon, Condorcet, Laplace, and Poisson. We still have not solved such problems in modern inferential statistics where we introduce a significance level that resembles the moral probability of those times.

With regards to the utility of money, the richer one is, the less it makes sense for the person to gain more money in this way. Daniel Bernoulli (1738/1954) suggested that the utility of money should be a logarithmic function. With this in mind, the expected utility of the game would be

$$E\big(u(X)\big) = E\big(\ln(X)\big) = E\left(\sum_{n=1}^{\infty} \ln\big(2^{n-1}\big)\frac{1}{2^n}\right) = \ln 2,$$

but this would not reflect the approach properly as with utility, the additional gain of the game would add to the riches r already owned. Thus an equation like

$$\ln(r) = E\big(\ln(r - s + X)\big) = E\left(\sum_{n=1}^{\infty} \ln\big(r - s + 2^{n-1}\big)\frac{1}{2^n}\right)$$

has to be solved. In this equation, s is the stake required to play the Petersburg game. If the relation above turns to an equation, then fair stakes are achieved. A numerical solution yields $s \approx 9$ for $r \approx 50{,}000$, which is quite close to that amount of money a lot of people would be prepared to stake in the game. Such a concept was called *moral expectation* but it depends on the riches at the beginning and on the utility function used.

To use utilities instead of money has been discussed at various periods and is part of economic models today and an element of decision theory but utilities are not seen as part of the foundations of probability. Some of the phenomena detected by Kahneman and Tversky (1979) also have their origin in that people do use utilities implicitly. However, the utility function differs between people and is very subjective. In newer attempts to cope with risks, utility is revived via the term impact. The search for an objective utility function has been in vain though it led to some empirical laws of decreasing impact like the Webner–Fechner law (see Stigler 1986) in physiology.

The second way is to introduce the idea of a moral probability that should be neglected. Several authors advocate differing rules for the size of such an entity. Buffon (1777) argued—in comparison to the probability of a 56 year old dying within 24 hours of $1/10{,}189$ (which he derived from mortality tables)—that probabilities smaller than 10^{-4} should be ignored as no one actually fears that. This would lead to neglecting all terms in the expectation with $n > 13$ and yield a corrected expected value of 6.5.

In the course of his likelihood argument for the solar system, Buffon required a rule of thumb for deciding when a specific likelihood is small enough to establish the 'proof' of a common cause. For practical purposes, he proposed that an event with a probability below 0.0001 should be considered impossible and its complement as certain. Such ambiguous rules about *moral certainty* can also be found with other writers like Cournot. More recently, Borel (1943) also argued that we should deal with small probabilities as if they were zero: A human being should neglect 10^{-6}, in the history of earth, 10^{-15} and, for the cosmos, 10^{-50} should be neglected. For a human being, terms with $n > 19$ would therefore be discarded and an expected value of 9.5 in the St Petersburg puzzle would result.

To develop the first "significance test" similar thoughts were applied by Arbuthnot (1710–1712) in his "proof" of a divine order of gender: the probability is very small that for 80 successive years more males than females are born, so the hypothesis that the gender proportion is equal has to be rejected as it would produce an observation with a probability of $(1/2)^{80} \approx 10^{-25}$ (Gigerenzer 2004, traces the early origins of statistical inference). Therefore, the alternative hypothesis must hold (probability for boys greater than for girls), which Arbuthnot interpreted as an expression of a divine order. Once we decide the limit for the moral probability, any hypothesis that gives the observed event a probability smaller than this limit is probabilistically *disproved*.

Bayes' Formula and Inverse Probabilities Bayes (1763) solved the following problem, which we will describe in modern terms—in its discrete version, the Bayesian formula as it is called now: Let $H_1, H_2, \ldots, H_k$ be k mutually exclusive and exhaustive statements (events) and $P(H_i)$ their (prior) probabilities. Further, let

E be an observable event, which has alternative probabilities relative to the statements, $P(E|H_i)$. Then—after E is observed, the (posterior) conditional probability of the statements H_i is given by

$$P(H_i|E) = \frac{P(H_i) \cdot P(E|H_i)}{\sum_{j=1}^{k} P(H_j) \cdot P(E|H_j)}.$$

To understand why this has been called the inverse problem, we go back to the special case that Bayes solved. The direct problem thereby relates to the development of the relative frequencies with increased number of trials

> [...] where the Causes of the Happening of an Event bear a fixed Ratio to those of its Failure, the Happenings must bear nearly the same Ratio to the Failures, if the Number of Trials be sufficient; and that the last Ratio approaches to the first indefinitely, as the number of Trials increases. (Hartley 1749, cited from Hald 2007, p. 25)

The indirect problem now relates to the development of the "causes" in the light of observed relative frequencies (cause might be better read as relative ease or probability and has nothing to do with cause in the original sense).

> An ingenious Friend [it is agreed widely that Hartley refers to Bayes] has communicated to me a Solution of the inverse Problem, [...] where the Number of Trials is very great, the Deviation must be inconsiderable: Which shews that we may hope to determine the Proportions, and, by degrees, the whole Nature, of unknown Causes, by a sufficient Observation of their Effects. (Hartley 1749, cited from Hald 2007, p. 25)

Bayes used a material situation to embody his model, in which his assumptions clearly make sense. His model includes a parameter p for an unknown probability and a series of data simulated on the device that follow a binomial distribution with parameters n and p. From these conditions, Bayes derived an integral term for the probability of p to lie within the limits of p_1 and p_2. In fact, he derived the posterior distribution of p after the event E of k successes in n trials of the Bernoulli experiment is observed (which is a beta distribution). His "prior" distribution on the parameter p as embodied in his material model was a uniform distribution. Independently of this situation, Bayes gave another *objective* reasoning for his uniform prior (see also Edwards 1978). If you do not know anything about the combined experiment (including the parameter p), then the stakes for any number k of successes in n trials should be the same. From this assumption it is relatively easy to justify a uniform prior for the unknown parameter p.

It is interesting to note that the main purpose of Bayes to clarify the relation between probabilities (p) and relative frequencies of an experiment performed under the same conditions was forgotten in the later discussion on the status of this prior: the form of the prior distribution loses its influence on the limiting behaviour as the posterior distribution concentrates more and more around the—previously unknown—point p and converges to it. This is the inverse side of Bernoulli's law of large numbers.

One of the mathematical conclusions from Bayes' formula was the rule of succession as discussed by Price (1763–1764, 1764–1765) in his introduction to Bayes' article. If one observes an event k times repeatedly in n successive observations, its probability equals $(k+1)/(n+2)$; this converges to 1 for $k=n$ as n tends to infinity. From the many times, the sun rose already, Buffon derives a moral certainty the sun will arise every (future) morning in his *Essai d'arithmétique morale* (1777).

As Bayes gave objective reasons for his uniform prior, it was justified as the prior as if it were inherent in reality and not a model about it. That clearly gave rise to speculations. On the one side, it was claimed in the argument that one does not know anything about the (combined) experiment and also not about the parameter p. On the other hand, a uniform distribution was derived for it. Thus, it seemed inconsistent from a philosophical point of view. How can complete ignorance on something lead to a concept that yields obviously some information about the parameter? To enhance the dilemma, we refer to the special case of $n=0$ and $k=0$ when there is neither data nor knowledge about any outcomes of trials, and the probability for the event occurring equals $1/2$ by the rule of succession.

The root of this trouble is the basic misunderstanding between objectivist and subjectivist concepts. Of course, with enough (independent) data the influence of the prior vanishes leaving no difference in the solution between frequentist and Bayesian (subjectivist) statisticians. If there is not enough data, then the question of using priors arises. Without a prior there is no solution at all; the alternative is to use a prior to get a solution. However, the prior used should really be based on an expert knowledge. Yet, it will never reach a status of objectivity.

4 Foundations and New Applications

4.1 Classical Probability

Laplace marks a culmination in the early conceptual development of probability but he already stands at the edge of a new era, which is marked by extensive use of continuous distributions either as an approximation to the binomial or independently as describing phenomena in physics and "social physics". Yet, philosophically his views are still based on a mechanical determinism and linked simply to ignorance. The Laplacean demon has become notorious:

> Given [...] an intelligence which would comprehend all [...] for it, nothing would be uncertain and the future, as the past, would be present to its eyes [...] Probability is relative, in part to this ignorance, in part to our knowledge. (1812/1951, p. 4, 6)

Laplace gave what is the first explicit definition of probability, the so-called classical probability, which underlies the classical a priori theory (APT). The probability $P(A)$ of an event A equals the ratio of the number of all outcomes which are favourable to event A to the number of all possible outcomes of the trial. This definition implicitly assumes that the individual outcomes are equally likely. Laplace varied the (correct but not well understood) justification for the prior distribution in the Bayes' formula into a general 'principle of insufficient reason' to underpin this assumption. According to this principle, we may assume outcomes are equally likely if we have no reason to believe that one or another outcome is more likely to arise. It is easy to see how this may arise out of gambling situations where the assumption would be seen as so obvious that it would not even be seen as an assumption, rather like Euclid's axioms.

This first formal definition had a great impact on the prestige of probability and boosted its applications and the development of techniques (e.g., generating functions, see Laplace 1814/1995) but did not clarify the nature of probability, as it refers to a philosophically obscure principle for its application and as its domain is far from the actual applications (continuous distributions) being studied at that time. Ensuing attempts to amend this principle have tried to replace this by indifference or invariance considerations with the aim to lend the equal probabilities an objective character under restricted conditions. From a modern viewpoint, Barnett (1973, p. 71) states that the principle "cannot stand up to critical scrutiny". There were also some problems and ideas which hindered conceptual progress.

Independence was a steady source of difficulty as it was not dealt with as a distinct concept. In a notorious example of a famous mathematician erring in probability, d'Alembert (1754) opposed the equi-probability of the four outcomes in tossing two coins; he argued fiercely in favour of the probabilities $1/3$ for each of the outcomes 0, 1, and 2 'heads'. While he used a primitive argument on three possible cases, his solution can be restored by using the succession rule and prior ignorance. Equally his solution can be refuted by reference to the independence concept. It is remarkable that writers always blamed d'Alembert for a wrong equi-probability argument but did not report on the dilemma between the subjectivist and objectivist interpretations of the problem, which lead to differing solutions.

Conversely, independence was used implicitly for combining arguments, which were not at all independent as in calculating the probability of the judgement of tribunals to be correct: the error probabilities of single judges were multiplied as if they were independent. If the tribunal consisted of 7 judges and each of them erred with probability 0.03, then the probability of the majority being right was (incorrectly) calculated to be 0.99997, which was judged as morally certain.

If Laplace's principle of insufficient reason is taken for granted, it would yield another ambiguous rule to transfer ignorance into knowledge. Thus, the focus in using it has to be on finding a representation of a problem, in which the possible cases may be viewed as equally likely. However, even physical symmetries can give rise to different embodiments, which leave the problem of attributing equal probabilities to these cases inconclusive. As a model, this induces no problem, as the basis for a mathematical theory, it does.

Indeed, Fine (1973, p. 167) illustrates the difficulties by the example of three models for two dice. While the first model is generally seen as the standard, the other models in fact are useful to describe phenomena in physics and the real world (or at least theories about it).

Maxwell–Boltzmann model: each of the 36 pairs (1, 1), ..., (1, 6), (2, 1), ..., (2, 6), ..., (5, 1), ..., (5, 6), (6, 1), ..., (6, 6), is equally likely; pairs like (2, 3) and (3, 2) are different outcomes in this model.

Bose–Einstein model: each of 21 pairs (1, 1), ..., (1, 6), (2, 2), ..., (2, 6), ..., (5, 5), (5, 6), (6, 6), is equally likely; pairs like (2, 3) and (3, 2) are treated as identical outcomes.

> *Fermi–Dirac model*: each of 15 pairs (1, 2), ..., (1, 6), (2, 3), ..., (2, 6), ..., (5, 6), is equally likely; the two components are forbidden to have the same value.

For ordinary dice, the Maxwell–Boltzmann approach is the usual and apparently natural model. The two dice can be (at least theoretically) discriminated; blue and red dice, or first and second trial; the independence assumption is accepted as highly plausible. This supposedly natural model, however, does not fit some applications in physics. According to Feller (1957, p. 39), numerous attempts have been made

> to prove that physical particles behave in accordance with Maxwell–Boltzmann statistics but modern theory has shown beyond doubt that this statistics does not apply to any known particles.

The behaviour of photons, nuclei and atoms containing an even number of elementary particles which are essentially indistinguishable may be best described by the Bose–Einstein statistics. Electrons, protons and neutrons where the particles are essentially indistinguishable and each state can only contain a single particle are well described by the Fermi–Dirac model. It is quite startling that the world is found to work in this way, experimentally. The situation is comparable to cases where non-Euclidean geometry is actually found to be a better model of the real world than the "natural" Euclidean geometry.

Symmetries are dependent on the purpose of modelling. It is salutary and interesting to note that pedagogical problems also arise over this issue. Therefore, probability should not be viewed as an inherent feature of real objects but only as a conceptual outcome of our endeavour to model reality. The intention to derive unique probabilities leads to various paradoxes. Fine (1973, p. 167) gives many counterexamples and summarizes his critique against Laplace's definition:

> We cannot extract information (a probability distribution) from ignorance; the principle is ambiguous and applications often result in inconsistencies; and the classical approach to probability is neither an objective nor an empirical theory.

4.2 Continuous Distributions

Continuous distributions are used for the first time by de Moivre (1738/1967). He investigates the probability of deviations of a binomial variable H_n from its expected value np:

$$P\big(|H_n - np| \le d\big)$$

He derives an approximation in terms of what we now call the normal distribution. For de Moivre the limiting expression served only as an approximation. Simpson (1755) in his attempt to give an argument for using the mean of observations instead of single observations used a triangular distribution. To describe error distributions, rectangular, triangular, quadratic distributions, and Laplace's double exponential distribution were used.

With his central limit theorem, Laplace (1810) paved the way for the normal distribution: the binomial distribution approaches the normal as the number of trials tends to infinity. Laplace immediately recognized its importance and formulated an intuitive rule when a normal distribution was to be expected: in analogy to the situation in the binomial case, which is built up by single summands of 1 or 0 (depending on the outcome of the like binomial trial), a variable should follow the normal law of probabilities if it is (virtually or really) built up by adding a large number of quantities. Following this interpretation, the normal distribution was soon the main distribution used in describing observational error or physical quantities.

There were two urgent problems arising from the more and more experimentally oriented sciences. The first was to give a probabilistic justification of the use of the method of least squares to derive an estimate for an unknown quantity from observations. The second was to give a probability argument for using the arithmetic mean instead of single observations, in order to "eliminate" measurement errors. Gauss (1809) uses the normal distribution in its own right and not only as an approximation within the context of error theory. His proof of the method of least squares was based on Laplace's inverse probability with a uniform prior on the location parameter and a normal distribution for the observations. Gauss also derived a functional equation from the condition that the mean is the best value—in the sense of the most probable value—to extract from a series of observations that lead to the normal distribution as solution. The concepts of maximum likelihood, method of least squares, and normal distribution are related and these interrelations supported their usage.

The new ideas were soon picked up by researchers, especially Laplace's intuitive argument was convincing: Bessel (1818) checked the validity of the normal distribution against empirical data of measurements. Quêtelet (1835) extended the idea of error distribution to biometric measurements developing the concept of l'homme moyen, an ideal figure that gets its value by the interference of a large sum of errors of nature. Thus, any biometric variable of human beings or animals should follow a normal distribution. The myth of the ubiquity of the normal distribution was born. Galton (1889) constructed the Quincunx to demonstrate how Nature would realize its elementary errors to add up to the final value of the variable under investigation. This apparatus is an elegant embodiment of the central limit theorem for the special case of a sum of binomial trials.

The enthusiasm with the normal distribution did not end with the systematic search of Pearson (1902) for systems of continuous distributions, nor with the Maxwell distribution for velocities in physics. Physics, especially the field of thermodynamics, saw a rapid development of concepts using the idea of distribution to describe the system's behaviour at the microscopic level in order to derive their laws of entropy on the macroscopic level. The developments in physics and especially in thermodynamics are seen as a driving force to develop a sound basis (see Batanero et al. 2005) and lead to Hilbert's (1900) programme to find an axiomatic basis for physics *and* probability. The more theoretical line of development was pursued by the Russian school (see Maistrov 1974): Following Poisson (1837), the Russians (Chebyshev 1887; Markov 1913) pursued various generalizations of the central limit theorem to derive the normal distribution of a sum under ever generalized conditions for the single parts of the sum.

4.3 Axioms of Probability

Various endeavours have been undertaken to develop a basis of the theory of probability following the idea of relative frequencies, culminating in the attempt by von Mises (1919) to give an axiomatic approach to the discipline based on idealized properties of random sequences, yet they were futile (Porter 1986) and amended only much later (Schnorr 1971). Since the turn of the century, the gold standard of a mathematical theory requires an axiomatic foundation. It is amazing that—despite Hilbert's (1900) agenda for the axiomatic foundation of physics and probability—it was not before 1933 that Kolmogorov was successful. This has to be contrasted with the great progress in measure theory around 1900. Lebesgue (1902) formulated the central problem of measure theory as the question of which sets can have a measure (as an extension of length, area, etc.). The problem was solved by the community using the class of Borel sets and the agreement (after some dispute) that the characteristic property of such measures should be captured by their sigma additivity. Borel (1909) applied this new knowledge to prove the strong law of large numbers and showed that the set of trajectories (of relative frequencies), which do not converge to the underlying probability has a zero measure. Such measures on sets of trajectories in three-dimensional space (which are elements of an infinite-dimensional space) have also become important to describe processes and laws in physics (especially in thermodynamics). What hindered the community to find a common logical base for this measure on the trajectories and the probabilities at the outset, which were still based on Laplace's equi-probability?

One obstacle surely was to find the common basis for discrete and continuous distributions. Another hindrance was the key concept of independence that is not contained in measure theory. Kolmogorov (1933), in his milestone publication, solved both obstacles. He came from almost innocent-looking axioms to the concept of distribution functions ($F_X(x) = P(X \leq x)$) with the result that there are certain types of distributions, including discrete and continuous distributions. For a succinct exposition of the mathematical ideas and concepts, see Borovcnik et al. (1991). Moreover, Kolmogorov dealt with independence as a key concept additional to the axioms. His approach is open to diverse interpretations of probability, including APT, FQT, and SJT, and has been taken as the mathematical foundation for probability.

Thus, on the one side, the prestige of the mathematical foundation let the applications boom again; on the other side, it laid the basis for a fierce controversy in the foundations starting with de Finetti (1937) who presented a system of axioms based on the interpretation of probability as degree of belief and ideal preference behaviour. Today, we can state that the battle remains unsettled and does not even clarify the various positions. However, the theories can be formulated in a way to conform the axioms of Kolmogorov. That is to say that most subjectivist followers of de Finetti would read the Kolmogorov axioms in the same way as objectivists but have their different interpretations. The schools would, however, differ considerably in their approach to statistical inference.

From the perspective of analytic science, Stegmüller (1973) agreed that statistical inference cannot be derived from the position of axiomatic probability alone but has to refer to further rationality criteria, which still have gaps. One unsolved problem is the difficulty to derive any direct statement for a single case from probabilistic information that is based on a long-run interpretation. There is no way to justify a significance level or power (of a test) or a confidence level (for confidence intervals) for the single application of a test. Popper's (1957) concept of propensity is just an elegant wording to the problem without a solution. According to that, probability is a physical property of an object or situation to produce events—a propensity.

From the view of applications, today the usage is liberal. If one has enough data then the use of methods of statistical inference without prior distributions on parameters leads to almost the same conclusions as subjectivist methods that make use of prior distributions, so there should be no dispute about the result even if the interpretation of it would still differ. If such frequentist information is lacking, the use of prior distributions has to be carefully calibrated against expert knowledge with the usual caveats in interpreting the final result as a guidance to decide but not as the final decision. What is difficult for probability is that the axiomatic foundation is not sufficient. The concepts cannot preserve their character once they are applied. For teaching, it means that the full comprehension of probability can neither be detached from an axiomatic perspective, nor from the inferential part of stochastics to preserve wider meaning.

5 Modern Philosophical Views on Probability

We now discuss the three main approaches to the nature of probability which are relevant for school mathematics. They are summarized below as classical a priori, frequentist, and subjectivist. The structural view is discussed as a sort of synthesis. We also present some virtues and some criticism of each approach; a fuller treatment can be found in Barnett (1973), or in Weatherford (1982), which also deals with other philosophical positions.

Classical a Priori Theory (APT) Following Laplace, the probability of a combined event is obtained by the fraction of outcomes favourable to this event in the sample space; this makes use of an implicit assumption of equal likelihood of all the single outcomes of the sample space. It is an a priori approach to probability in that it allows calculation of probabilities *before* any trial is made. Geometric probability is closely related to it; it reduces the probability concept to counting or area.

In the case of applications, one is confronted with the problem of deciding the single outcomes that are equally likely, leading to the charge of circularity. Symmetry in the physical experiment with respect to Laplace's 'principle of insufficient reason' is a shaky guideline to help in this respect. One major philosophical problem is that the same physical experiment can reveal several different symmetries (as for the dice above). Hence the criticism is that the logic of symmetry is useless for prediction. Another criticism is that APT is subjective with regards to choosing the underlying symmetry. Furthermore, it is not clear how APT should be adapted by experience—for example, in the case of a (biased) die which lands on 3 for ten consecutive tosses. Conversely, there is virtually universal agreement that the probability of a die (with no obvious bias) to land on 6 is $1/6$. There are children who deny this from their experience of board games, but this is linked to expectation (of waiting times till the first six) rather than to the probability of getting sixes. As we have seen, such examples from gambling were the source of the fundamental idea of probability. There was no intention that probability should apply to actuarial tables; for some it still seems odd to calculate the chance of dying at different ages, based on mortality tables. Thus some would say that there are two fundamental notions of probability: one is based on symmetries whilst the other is based on data.

Frequentist Theory (FQT) The probability of an event is obtained from the observed relative frequency of that event in repeated trials. Probabilities are never obtained exactly by this procedure but are estimated. It is an a posteriori, experimental approach based on information *after* actual trials have been done. The measure of uncertainty is assigned to an individual event by embedding it into a collective—an infinite class of 'similar' events which are assumed to have certain 'randomness' properties: these ideas were developed by von Mises. Then the probability is the limit towards which the relative frequency tends.

In applying this definition, one is faced with the problem that an individual event can be embedded in different collectives, with no guarantee of the same resulting limiting frequencies: one requires a procedure to justify the choice of a particular embedding sequence. Furthermore there are obvious difficulties in defining what is meant by 'similar' or by 'randomness'; indeed an element of circularity is involved. Even the notion of settling down presents difficulties in terms of the number of trials needed in long term frequency. More fundamentally, there are many events where a probability is required but it is not possible to carry out repeated experiments—this is especially true for events with a low probability. There are three main criticisms of FQT. The first is that FQT probabilities can, in principle, never be calculated. The second is that it cannot be known whether FQT probabilities actually exist. Finally, a tentative FQT value can never be confirmed or even denied. On the other hand, FQT probability is more suited to modern views and seems to be based on evidence rather than external thought. FQT also offers a practical means of calculating probabilities, especially in situations, such as mortality at different ages, where no symmetry can be applied. In some ways, FQT is seen as more scientific and certainly features in modern physics and biology.

Subjectivist Theory (SJT) Probabilities are evaluations of situations which are inherent in the individual person's mind—not features of the real world around us which is implicitly assumed in the first two approaches above. The basic assumption here is that individuals have their own probabilities which are derived from an implicit preference pattern between decisions. Ideas are taken from gambling where given any event, its probability can be determined by the odds a person is prepared to accept in betting on its occurrence. Obviously people may differ in the odds they would accept, but this does not matter provided basic rules of coherence and consistency are met. For example, it would be foolish to place two bets of 3 to 2 on both horses in a two-horse race (i.e. for a stake of £2 you could win £3) because one is bound to lose money as the win of £3 does not compensate for the loss of £4 overall. Coherence formalises this basic idea from which one can deduce the basic laws of probability (Barnett 1973). These ideas were developed in detail by de Finetti (1937), who starts his lifetime work (de Finetti 1974) by the startling claim that 'PROBABILITY DOES NOT EXIST'. His followers whilst being very vociferous, remain in a minority amongst mathematicians.

For a subjectivist there are two categories of information, namely prior information which is independent of any empirical data in a person's mind, and empirical data from frequencies in repeated experiments. Both types of information are combined in Bayes' formula to give a new probability of the event in question. This updating of probabilities is called 'learning from experience'. Bayes' formula, combined with simple criteria of rationality, allows a *direct* link from experience and prior information to decisions. Thus, the usual problems of statistical methods based on objectivist probability are circumvented.

A problem inherent to the subjectivist approach is its intended ubiquity; any uncertainty has to be specified by probabilities. There might be occasions when it is better not to force meagre information into the detailed form of a distribution; it might be wiser to try a completely different method to solve the problem. The most striking argument against the subjectivist position, however, is that it gives no guidance on how to measure prior probabilities (Barnett 1973, p. 227). Though there are flaws in the classical and frequentist approaches, they provide pragmatic and accepted procedures to calculate probabilities. Of course, a subjectivist could exploit all frequentist and symmetry information to end up with a suitable prior distribution; but there is nothing in the theory to encourage such an approach. Indeed any probability can be assigned to any event by an individual; the only constraint is to avoid incoherence such as described above for a two-horse race.

The initial prior probabilities can be changed as new information is obtained and this can be done using Bayes' theorem: this leads to the name of Bayesians for subjectivists, but Bayes' theorem is equally applicable in APT and FQT. The main criticism of SJT is its intrinsic subjectivity and lack of an objective basis: it seems to confuse feeling with fact. This implies that it does not matter what we believe and whether there is any evidence to the contrary, apart from the relatively less significant proviso to lack coherence. However, SJT is valid for cases where there is no symmetry for APT or evidence from repeated trials as required by FQT. For example, one may wish to discuss the probability that an incumbent president will win the next election. One may have a range of information about such an event and arrive at a probability; such an approach is certainly taken in the media. Indeed, it is often the case that such an approach is necessary in business decisions where knowledge is partial and yet a decision has to be made. One only has to think about the way that ideas of risk have become more prominent in the last decade, with the financial crisis and other events such as the spread of epidemics across the world. The SJT is ideally suited to such scenarios.

Commentary In the structural approach, typically taken in courses at higher levels, formal probability is a concept which is implicitly defined by a system of axioms and the body of definitions and theorems which may be derived from these axioms. Probabilities are derived from other probabilities according to mathematical theorems, with no justification for their numerical values in any case of application. This structural approach does not clarify the nature of probability itself though the theorems derived are an indicator of possible interpretations. The structural approach serves as a theoretical framework for the two main conceptions of probability.

The *objectivist* position encompasses the classical and the frequentist view; according to it, probability is a kind of disposition of certain physical systems which is indirectly related to empirical frequencies. This relation is confirmed by theorems like the law of large numbers. The *subjectivist* view treats probability as a degree of confidence in uncertain propositions (events). Axioms on rational betting behaviour like coherence and consistency provide rules for probabilities; the Kolmogorov axioms prove mathematical theorems which must be obeyed if one wants to deal rationally with probabilities.

Historically, physics was dominated by a causal approach, which also left its traces in the philosophy of randomness as may be seen by the Laplacean demon, which turned probability to the tool of those who are ignorant. More recently, however, this causal approach in physics has switched to a completely random approach where everything is referred back to random laws (see Styer 2000). Remarkably, sometimes the discussants forget about the use of a model in this approach and perhaps infer that the whole world is random. However, there are alternative approaches in the foundations of physics, which stress causality (e.g. Dürr et al. 2004). The causality—randomness debate will surely continue.

Within the connections between randomness and causality in physics, the controversy in the foundations of probability, between those who promote an objective character of probability (the Objectivists) and those who believe that probability is fundamentally subjective (the Bayesians), was fiercely fought but has been decided in favour of an interpretation of Kolmogorov's axiomatic probability as something like a propensity in the sense of Popper (1957). Such a decision was taken in analytic philosophy despite the gaps in rational argument of statistical inference based on probability. Stegmüller (1973) asserted that a radical subjective physics cannot be accepted and, to accept the gap in rationality, is less problematic. However, apart from physics, the Bayesian approach has been widely accepted in the sense of Berger (1993). If enough data is available, then a classical (objectivist) approach may be used, if not then qualitative information has to be used and the problem remains how to make it more communicable between those who use the model.

For teaching, the discussion in the American Statistician in 1997 has clearly shown some disadvantages of a classical approach for statistical inference (Albert 1997; Berry 1997; and Witmer et al. 1997). But as Moore (1997) states, Bayesian methods are too complicated to teach initially. In fact, Bayesian methods for statistical inference are easier to understand but—as they deal with qualitative data—much harder to apply, at least when frequentist information is missing. Vancsó (2009) devotes special attention to exploit the advantages of both approaches simultaneously in teaching.

This shows, on the one hand, the close connection of probability to physics, which will become even stronger as physics might abandon the causal paradigm in favour of a complete randomization ideal. On the other hand, statistical inference has evolved greatly by the sound foundation of probability in the last century, and revived the subjective roots of probability again. As it is not possible to close the gaps in rationality, the community follows the same pattern of reaction as in the case of Gödel's fundamental critique to mathematics, namely to use that approach which is more suitable in each situation without being able to universalize the foundations.

Finally, we turn to topical aspects of probability relating to risk. As may be seen from the historic development, risk has always played a prominent role in the development of the concepts. This is not restricted to games of chance but extends to more general situations under uncertainty. By risk usually we refer to an outcome, which has a high negative impact, which should be avoided. More neutrally, in some situations, some outcomes have a small impact or may have a benefit. The decision situation involves several options, all with different outcomes related to the impact (loss or win). Which of the decisions is best? For a perceptive discussion of situations involving risks in law or medicine, see Gigerenzer (2002).

The difficulties are increased by the fact that many such situations involve very small probabilities (e.g. the prevalence of *BSE*) but the potential impact is extremely negative. The other difficulty is to separate the awareness of the probabilities and their measurement from the perception of the impact. High (negative or positive) impact might get so dominant that the probabilities associated to it may be "forgotten". Moreover, as the probabilities are often so small, they cannot be measured by frequencies. This links to Bernoulli's discussion of how to combine stochastic arguments.

These topics pose a challenge, not only for teaching but also for further steps of conceptual development. For teaching, the challenge is discussed in Kapadia and Borovcnik (1991), Jones (2005), Borovcnik and Kapadia (2009), and Batanero et al. (2011). More recently, the perspective of modelling with multiple models in a problem situation has been explored by Borovcnik and Kapadia (2011).

6 Concluding Comments

The key results in probability theory are similar in both FQT and SJT, as well as in APT when it applies. At least this is true if one neglects the more radical subjectivist variants like de Finetti, who denies even the σ-additivity of probability measures. That means that the competing positions use the same axiomatic language with respect to syntax. Of course, the semantics are different; Kolmogorov's axioms are not the basic medium for the interpretation between theory and reality, but are logically derived rules. Despite the formal equivalence of the approaches, Kolmogorov's axioms are usually thought to be a justification of the objectivist, especially the frequentist view.

To illustrate the various philosophical positions, we present a standard example of tossing a particular blue die in which APT, FQT, and SJT can be used. The classical view would assign a probability of $1/6$ of getting a six as there are six faces which one can assume are all equally likely; the frequentist view would assign a probability by doing repeated experiments or, as experiments with other dice have yielded $1/6$, the same value would be assigned for the die in question. In both these viewpoints, the probability is seen as an inherent feature of the die, and the tossing procedure. The subjectivist would assert that the probability was a mental construct which he/she may alter if new information became available about the die. For example, a value different to $1/6$ might be assigned if the die is black or heavier or smaller or was slightly chipped. There ought to be some basis for the decision made, but this is not a requirement for de Finetti. However, since the probability is not viewed as inherent in the object, the conflicting values would not involve any logical problems. In fact, the subjectivist does not reject considerations of symmetry or frequency—both are important in evaluating probabilities. There is an expectation, however, that ideas of symmetry or frequency are made explicit if applied.

In any case of application, one has to choose a specific subjective or objective interpretation to determine the model and inherent probability. Philosophically, for the structuralist, there is no way of deciding which model is better. This controversy on the nature of probability will continue. Fine (1973) states that

> subjective probability holds the best position with respect to the value of probability conclusions, [...]. Unfortunately, the measurement problem in subjective probability is sizeable and conceivably insurmountable [...]. The conflict between human capabilities and the norms of subjective probability often makes the measurement of subjective probability very difficult. (p. 240)

This extensive historical and philosophical background provides a valuable perspective for teachers. It is for them to decide how to present ideas to pupils and students, whether by APT, FQT, or SJT. We assert that a judicious combination of all three approaches is required to build up firm concepts from emerging intuitions as has happened in history, leading to the fiercely competing philosophical positions.

References

Albert, J. (1997). Teaching Bayes' rule: a data-oriented approach. *The American Statistician*, *51*(3), 247–253.

Arbuthnot, J. (1710–1712). An argument for divine providence taken from the constant regularity observed in the birth of both sexes. *Philosophical Transactions of the Royal Society*, *27*, 186–190. Reprinted in M. Kendall, & R. L. Plackett (Eds.) (1977), *Studies in the history of statistics and probability* (Vol. 2, pp. 30–34). London: Griffin.

Barnett, V. (1973). *Comparative statistical inference*. New York: Wiley.

Batanero, C., Henry, M., & Parzysz, B. (2005). The nature of chance and probability. In A. G. Jones (Ed.), *Mathematics education library: Vol. 40. Exploring probability in school: challenges for teaching and learning* (pp. 15–37). New York: Springer.

Batanero, C., Burrill, G., & Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI/IASE study: the 18th study*. New York: Springer.

Bayes, T. (1763). An essay towards solving a problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society*, *53*, 370–418. Reprinted in E. S. Pearson, & M. G. Kendall (1970), *Studies in the history of statistics and probability* (Vol. 1, pp. 131–154). London: Griffin.

Bellhouse, D. R. (2000). De Vetula: a medieval manuscript containing probability calculations. *International Statistical Review*, *68*(2), 123–136.

Berger, J. O. (1993). *Statistical decision theory and Bayesian analysis*. New York: Springer.

Bernoulli, D. (1738/1954). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *5*, 175–192. English translation: Sommer, L., Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23–36.

Bernoulli, J. (1713/1987). *Ars conjectandi*. Basel: Impensis Thurnisiorum, Fratrun. Translation of 4th part by N. Meunier. Rouen: Institut de Recherche sur l'Enseignement Mathematique; German translation by R. Haussner (1899), *Wahrscheinlichkeitsrechnung*. Leipzig: Engelmann.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician*, *51*(3), 241–246.

Bessel, F. W. (1818). Untersuchungen über die Wahrscheinlichkeit der Beobachtungsfehler. *Astronomische Nachrichten*, *15*, 358–359, 368–404.

Borel, É. (1909). Les probabilités dénombrables et leurs applications arithmétique. *Rendiconti del Circolo Matematico di Palermo*, *27*(1), 247–271.

Borel, É. (1943). *Les probabilités et la vie*. Paris: Presses Universitaires de France.

Borovcnik, M. (2011). Strengthening the role of probability within statistics curricula. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education. A joint ICMI/IASE study: the 18th study* (pp. 71–83). New York: Springer.

Borovcnik, M., & Kapadia, R. (Eds.) (2009). Research and developments in probability education. *International Electronic Journal of Mathematics Education*, *4*(3), 111–130. Online: www.iejme.com/032009/main.htm.

Borovcnik, M., & Kapadia, R. (2011). Modelling in probability and statistics—key ideas and innovative examples. In J. Maaß, & J. O'Donoghue (Eds.), *Real-world problems for secondary school students—case studies* (pp. 1–44). Rotterdam: Sense Publishers.

Borovcnik, M., Bentz, H.-J., & Kapadia, R. (1991). A probabilistic perspective. In R. Kapadia, & M. Borovcnik (Eds.), *Mathematics education library: Vol. 12. Chance encounters* (pp. 27–71). Dordrecht: Kluwer Academic.

Buffon, G. L. (1777). Essai d'arithmetique morale. In G. L. Buffon, *Histoire naturelle générale et particulière* (Suppl. 4). Paris: Imprimérie Royale. English translation (1785), *Natural history, general and particular* (2nd ed.). London: Strahan and Cadell.

Cardano, G. (approx. 1564). Liber de ludo aleae, published 1663 in *Opera Omnia*, *1*. Lyon: Sponius.

Chebyshev, P. L. (1887). Sur deux théoremes relatifs aux probabilité. *Acta Arithmetica*, *XIV*, 805–815. French transl. 1890–1891.

Çinlar, E. (2011). *Probability and stochastics*. Berlin: Springer.

d'Alembert, J. (1754). Croie ou pile. In D. Diderot, & J. d'Alembert, *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (Vol. 4, pp. 512–513). Paris: Société des Gens de Lettres.

Daston, L. (1988). *Classical probability in the enlightenment*. Princeton: Princeton University Press.

David, F. N. (1962). *Games, gods and gambling*. London: Griffin.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales Institut Henri Poincaré*, *7*, 1–68. Foresight: its logical laws, its subjective sources (Chaps. 1–4 translated by H. E. Kyburg Jr.). In S. Kotz, & N. L. Johnson (1992), *Breakthroughs in statistics. Vol. I. Foundations and basic theory* (pp. 134–174). New York: Springer.

de Finetti, B. (1974). *Theory of probability*. New York: Wiley. Translated by A. Machi, & A. Smith.

de Moivre, A. (1738/1967). *The doctrine of chances* (2nd ed., fuller, clearer, and more correct than the first). London: Woodfall. Reprint of 3rd ed. 1967. New York: Chelsea. 1st ed. 1718, London: Pearson.

Dürr, D., Goldstein, S., Tumulka, R., & Zanghi, N. (2004, August, 24). Bohmian mechanics and quantum field theory. *Physical Review Letters*, *93*(9), 090402. Online: http://arXiv.org/abs/quant-ph/0303156.

Edwards, A. W. F. (1978). Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics*, *5*, 116–118.

Feller, W. (1957). *An introduction to probability theory and its applications* (2nd ed.). New York: Wiley.

Fermat, P., & Pascal, B. (1679). Correspondence 1654. In P. Fermat: *Varia opera mathematica*. Toulouse: Joannem Pech. English translation in F. N. David (1962), *Games, gods and gambling* (pp. 229–253). London: Griffin.

Fine, T. L. (1973). *Theories of probability*. New York: Academic Press.

Galton, F. (1889). *Natural inheritance*. London: Macmillan.

Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Hamburg: Perthes und Besser.

Gigerenzer, G. (2002). *Calculated risks: how to know when numbers deceive you*. New York: Simon & Schuster.

Gigerenzer, G. (2004). Die Evolution des statistischen Denkens. *Stochastik in der Schule*, *24*(2), 2–13.

Graunt, J. (1662). Natural and political observations upon the Bills of Mortality, chiefly with reference to the government, religion, trade, growth, air, diseases etc. of the City of London. *Royal Society of London*.

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

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Hald, A. (1990). *A history of probability and statistics and their applications before 1750*. New York: Wiley.

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.

Hartley, D. (1749). *Observations on man, his frame, his duty, and his expectations*. London: Richardson. Reprinted 1996 with an introduction by T. L. Huguelet. Gainesville: Scholar's Fascimiles and Reprints.

Hilbert, D. (1900). Mathematische Behandlung der Axiome der Physik. *Akademie der Wissenschaften in Göttingen*, *1900*, 272–273. Online: http://aleph0.clarku.edu/~djoyce/hilbert/problems.html#prob6 (English translation). From the lecture "Mathematische probleme" delivered at the International Congress of Mathematicians in Paris 1900. Online: www.mathematik.uni-bielefeld.de/~kersten/hilbert/problems.html.

Huygens, C. (1657). De ratiociniis in ludo aleae. In F. v. Schooten: *Exercitationes matematicae*, Leyden: Elsevirii.

Jones, G. A. (2005). *Mathematics education library: Vol. 40. Exploring probability in schools. Challenges for teaching and learning*. New York: Springer.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *XLVII*, 263–291.

Kapadia, R., & Borovcnik, M. (Eds.) (1991). *Mathematics education library: Vol. 12. Chance encounters*. Dordrecht: Kluwer Academic.

Kolmogorov, A. N. (1933). *Ergebnisse der Mathematik. 2. Band, Heft 3. Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. Reprinted 1977 (62 pp.). English translation 1956: *Foundations of the theory of probability*. New York: Chelsea.

Laplace, P. S. (1810). Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités. *Mémoires de L'Académie des Sciences Paris*, *1809*, 353–415, 559–565.

Laplace, P. S. (1812/1951). Essai philosophique sur les probabilités. *Journal de l'École Polytechnique*, *VII/VIII*, 140–172. English translation: *A philosophical essay on probabilities* (extended version). New York: Dover.

Laplace, P. S. (1814/1995). *Théorie analytique des probabilités* (2nd ed.). Paris: Courcier. 1st ed. 1812. Reprinted 1995, Paris: Jacques Gabay.

Lebesgue, H. (1902). Intégrale, Longeur, Aire. Thése. *Annali di Matematica Pura ed Applicata*, *7*(3), 231–359. Reprinted in H. Lebesgue (1972/1973), Oeuvres scientifiques (Vol. 1, pp. 203–331). Genève: Université de Genève, Institut de Mathématiques.

Logos Bible Software (n.d.). *The Holy Bible*. English standard version. Containing old and new testaments. Online. http://biblia.com/.

Maistrov, L. E. (1974). *Probability theory. A historical sketch*. New York: Academic Press.

Markov, A. A. (1913). *A probabilistic limit theorem for the cases of academician A. M. Lyapunov*. Reprinted in A. A. Markov (1951), Selected works of A. A. Markov in the theory of numbers and the theory of probability (in Russian). Leningrad: Izdatel'stvo Akademii Nauk SSSR, 319–338.

Moore, D. S. (1997). Bayes for beginners? Some reasons to hesitate. *The American Statistician*, *51*(3), 254–261.

Ore, Ø. (1953). *Cardano: the gambling scholar. With a translation from Latin of Cardano's "Book on games of chance" by S. H. Gould*. Princeton: Princeton University Press. Reprinted 1965, New York: Dover.

Pacioli, L. (1494). *Summa de arithmetica, geometria, proportioni et proportionalitá*. Venedig: Paganino de' Paganini.

Pearson, K. (1902). On the systematic fitting of curves to observations and measurements—Part I. *Biometrika*, *1*, 265–303.

Peverone, G. F. (1558). *Due brevi e facili trattati, il primo d'arithmetica, l'altro di geometria*. Lione: G. di Tournes.

Poisson, S. D. (1837). *Recherches sur la probabilité des judgements en matiére criminelle et en matiére civile*. Paris: Bachelier.

Popper, K. R. (1957). The propensity interpretation of the calculus of probability and the quantum theory. In S. Körner (Ed.), *Observation and interpretation in the philosophy of physics*. London: Butterworth.

Porter, T. (1986). *The rise of statistical thinking, 1820–1900*. Princeton: Princeton University Press.

Price, R. (1763–1764). Introduction and appendix to Bayes' Essay. *Philosophical Transactions of the Royal Society of London*, *53*, 370–375; *54*, 404–418

Price, R. (1764–1765). A demonstration of the second rule in Bayes' Essay. *Philosophical Transactions of the Royal Society of London*, *54*, 296–297; *55*, 310–325.

Quêtelet, A. (1835). *Sur l'homme et le dévelopement de ses facultés, ou Essai de Physique Sociale*. Paris. Online: California Digital Library. www.archive.org/details/surlhommeetled00quet.

Schneider, I. (1989). *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933—Einführung und Texte*. Berlin: Akademie-Verlag.

Schnorr, C. P. (1971). *Lecture notes in mathematics: Vol. 218. Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Berlin: Springer.

Shafer, G. (1996). The significance of Jacob Bernoulli's Ars conjectandi for the philosophy of probability today. *Journal of Econometrics*, *75*, 15–32. Online: www.glennshafer.com/cv.html#articles (Vol. 55, preprint version), or www.sciencedirect.com/science/journal/03044076/75/1.

Simpson, T. (1755). A Letter to the President of the Royal Society, on the advantage of taking the mean of a number of observations in practical astronomy. *Philosophical Transactions of the Royal Society, Series A*, *49*, 82–93. Online: available at Google books.

Stegmüller, W. (1973). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie* (Vol. 4). Berlin: Springer. First part: Personelle Wahrscheinlichkeit und Rationale Entscheidung. Second part: Personelle und statistische Wahrscheinlichkeit.

Stigler, S. M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Styer, D. F. (2000). *The strange world of quantum mechanics*. Cambridge: Cambridge University Press.

Todhunter, I. (1865). *A history of the mathematical theory of probability from the time of Pascal to that of Laplace*. Cambridge and London: Macmillan & Co. Online: California Digital Library. www.archive.org/details/ofmathemahistory00todhrich.

Vancsó, Ö. (2009). Parallel discussion of classical and Bayesian ways as an introduction to statistical inference. *International Electronic Journal in Mathematics Education*, *4*(3), 291–322. Online: www.iejme.com/032009/main.htm.

Venn, J. (1888). *The logic of chance* (3rd ed.). London: Macmillan. 1st ed. published 1866.

von Mises, R. (1919). Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, *5*, 52–99.

von Plato, J. (1994). *Creating modern probability: its mathematics, physics and philosophy in historical perspective*. Cambridge: Cambridge University Press.

Weatherford, R. (1982). *Philosophical foundations of probability theory*. London: Routledge & Kegan Paul.

Witmer, J., Short, T. H., Lindley, D. V., Freedman, D. A., & Scheaffer, R. L. (1997). Teacher's corner. Discussion of papers by Berry, Albert, and Moore, with replies from the authors. *The American Statistician*, *51*(3), 262–274.

Springer
http://www.springer.com/978-94-007-7154-3

Probabilistic Thinking
Presenting Plural Perspectives
Chernoff, E.J.; Sriraman, B. (Eds.)
2014, XXI, 747 p. 164 illus., Hardcover
ISBN: 978-94-007-7154-3