Acknowledgements I would like to thank Shmuel Elitzur for saving me from serious errors in the accounts of physical theories in this paper. Thanks also to the editors of this volume for very helpful criticisms of earlier drafts. This work was written with the aid of a grant from the Israel Science Foundation.

References


16

Probability—A Philosophical Overview

Alan Hájek
Professor of Philosophy
Australian National University

From the Editors

The work in the philosophy of probability seems much closer to the mathematical content than most work by philosophers of mathematics. This could be because probability is a relatively recent addition to the set of mathematical subjects, or maybe because of its origins in topics such as gambling. Perhaps because everyone has studied at least some mathematics in school, but not everyone has studied much probability, there are fewer philosophers working in the philosophy of probability. Alan Hájek is a philosopher with a deep interest in the philosophy of probability. His chapter is an introduction to many of the issues currently being discussed in the philosophy of probability. It should be accessible to anyone who has taken the standard undergraduate probability and statistics course. Neither of the editors of this volume have done any work in the field, but the questions here seem very natural to us.

Alan Hájek is a Professor of Philosophy in the Research School of Social Sciences at the Australian National University (philres.ansu.edu.au/people-defaults/alanh/index.php). His research interests include the philosophical foundations of probability and decision theory, epistemology, the philosophy of science, metaphysics, and the philosophy of religion. His paper "What Conditional Probability Could Not Be" won the 2004 American Philosophical Association Article Prize for the best article published in the previous two years by a younger scholar. The Philosopher's Annual selected his "Waging War on Pascal's Wager" as one of the ten best articles in philosophy in 2003. Other articles of interest are "Perplexing Expectations" (with Harris Nover), Mind 115, (2006) and "The Cubic Guy Paradox", Analysis, 65: 2, (2005). He has a book in preparation, Arrows and Haloes: Probabilities, Conditionals, Desires, Beliefs to be published by Oxford University Press.
1 Personal and Pedagogical Prologue

Once upon a time I was an undergraduate majoring in mathematics and statistics. I attended many lectures on probability theory, and my lecturers taught me many nice theorems involving probability: ‘$P$ of this equals $P$ of that’, and so on. One day I approached one of them after a lecture and asked him: “What is this ‘$P$’ that you keep on writing on the blackboard? What is probability?” He looked at me like I needed medication, and he told me to go to the philosophy department. In the interests of pedagogy, in retrospect I think that he could have benefited from some discussions with philosophers. For when I now teach those same theorems to my students, I hope that I can imbue them with deeper meaning and motivation when I point out what is at stake philosophically.

Anyway, I did go to the philosophy department. (Admittedly, my route there was long and circuitous.) There I found a number of philosophers asking the very same question: what is probability? All these years later, it’s still one of the main questions that I am working on. I still don’t feel that I have a completely satisfactory answer, although I like to think that I’ve made some progress on it. For starters, I know many things that probability is not, namely various highly influential analyses of it that cannot be right—we will look at them shortly. As to promising directions regarding what probability is, I will offer my best bets at the end, concluding with some further personal and pedagogical thoughts.

2 Introduction

Bishop Butler’s dictum [Butler 1736] that “Probability is the very guide of life” is as true today as it was when he wrote it in 1736. It is hardly necessary to point out the importance of probability in statistics, physics, biology, chemistry, computer science, medicine, law, meteorology, psychology, economics, and so on. Probability is crucial to any discipline that deals with indeterministic processes, any discipline in which evidence has a non-deductive bearing on hypotheses, indeed any discipline in which our ability to predict outcomes is imperfect—which is to say virtually any serious empirical discipline. Probability is also seemingly ubiquitous outside the academy. Probabilistic judgments of the efficacy and side-effects of a pharmaceutical drug determine whether or not it is approved for release to the public. The fate of a defendant on trial for murder hinges on the jurors’ opinions about the probabilistic weight of evidence. Geologists calculate the probability that an earthquake of a certain intensity will hit a given city, and engineers accordingly build skyscrapers with specified probabilities of withstanding such earthquakes. Probability undergirds even measurement itself, since the error bounds that accompany measurements are essentially probabilistic confidence intervals. We find probability wherever we find uncertainty—that is, almost everywhere in our lives.

It is surprising, then, that probability is a comparative latecomer on the intellectual scene. To be sure, inklusive ideas about chance date back to antiquity—Epicurus, and later Lucretius, believed that atoms occasionally underwent indeterministic swerves. In the middle ages, Averroes had a notion of ‘equipotency’ that might be regarded as a precursor to probabilistic notions. But probability theory was not conceived until the 17th century, when the study of gambling games motivated the first serious mathematical study of chance by Pascal and Fermat in the mid-17th century, culminating in the Port-Royal Logic. Over the next three centuries, the theory was developed by such authors as Huygens, Bernoulli, Bayes, Laplace, Condorcet, de Moivre, Venn, Johnson, and Keynes. Arguably, the crowning achievement was Kolmogorov’s axiomatization in 1933, which put probability on rigorous mathematical footing.

When I asked my professor “What is probability?”, there are two ways to understand that question, and thus two kinds of answer that could be given (apart from bemused advice to seek attention from a doctor, or at least a doctor of philosophy). First, the question may mean: what are the formal features of probability? That is a mathematical question, to which Kolmogorov’s axiomatization is the widely (though not universally) agreed upon answer. I review this answer in the next section as it was given to me at great length in my undergraduate statistics courses. Second, the question may mean: what sorts of things are probabilities—what, that is, is the subject matter of probability theory? This is a philosophical question, and while the mathematical theory of probability certainly bears on it, the answer must come from elsewhere—in my case, from the philosophy department.

3 The Formal Theory of Probability

3.1 Unconditional Probability

Kolmogorov begins his classic book ([Kolmogorov 1933]) with what he calls the “elementary theory of probability”: the part of the theory that applies when there are only finitely many events in question. Let $\Omega$ be a set (the ‘universal set’). A field on $\Omega$ is a set of subsets of $\Omega$ that has $\Omega$ as a member, and that is closed under complementation (with respect to $\Omega$) and finite union. Let $\Omega$ be given, and let $\mathcal{F}$ be a field on $\Omega$. Kolmogorov’s axioms constrain the possible assignments of numbers, thought of as probabilities, to the members of $\mathcal{F}$. Let $P$ be a function from $\mathcal{F}$ to the real numbers obeying:

1. (Non-negativity) $P(A) \geq 0$ for all $A \in \mathcal{F}$.
2. (Normalization) $P(\Omega) = 1$.
3. (Finite additivity) $P(A \cup B) = P(A) + P(B)$ for all $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$.

Such a triple $(\Omega, \mathcal{F}, P)$ is called a probability space.

Here the arguments of the probability function are sets, often called events. (Note that this is a technical sense of the term that may not neatly align with ordinary usage—for example, it is not clear that ‘events’ in the latter sense have the required closure properties.) Kolmogorov’s probability theory is thus dependent on set theory.

We could instead attach real-valued probabilities to members of a collection $S$ of sentences of a language, closed under finite truth-functional combinations, with the following counterpart axiomatization:

I. $P(A) \geq 0$ for $A \in S$.
II. If $T$ is a tautology, then $P(T) = 1$.
III. $P(A \lor B) = P(A) + P(B)$ for all $A \in S$ and $B \in S$ such that $A$ and $B$ are logically incompatible.

Note how these axioms take the notions of ‘tautology’, ‘logical incompatibility’ and ‘implication’ as already understood. To this extent we may regard probability theory, so formulated, as dependent on deductive logic.
Now let $\Omega$ be infinite. A non-empty collection $\mathcal{F}$ of subsets of $\Omega$ is called a sigma algebra (or sigma field, or Borel field) on $\Omega$ iff $\mathcal{F}$ is closed under complementation and countable union, i.e.

$$A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}.$$ 

Kolmogorov introduces a further ‘infinitary’ axiom.

4. (Continuity) If $E_1, E_2, \ldots$ is a sequence of sets such that $E_i \supseteq E_{i+1}$ and $\bigcap_{n=1}^{\infty} E_n = \emptyset$ then $P(E_n) \to 0$ (where $E_n \in \mathcal{F}$ for all $n$).

That is, if $E_1, E_2, \ldots$ is a sequence of non-increasing sets (according to the set-inclusion relation), with empty infinite intersection, then $\lim_{n \to \infty} P(E_n) = P\left(\bigcap_{n=1}^{\infty} E_n\right)$. Now, define a probability measure $P(-)$ on $\mathcal{F}$ as a function from $\mathcal{F}$ to $[0, 1]$ satisfying axioms 1–3, as before, and also the new axiom 4.

Equivalently, we can replace the conjunction of axioms 3 and 4 with a single axiom:

3*. (Countable additivity) If $\{A_i\}$ is a countable collection of (pairwise) disjoint sets, each $A_i \in \mathcal{F}$, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Thanks to the assumption that $\mathcal{F}$ is a sigma algebra, we are guaranteed that the probability on the left hand side is defined.

De Finetti ([de Finetti 1972] and [de Finetti 1974]) marshals a battery of arguments against countable additivity, most of them variations on these:

The infinite lottery: Suppose a positive integer is selected at random—we might think of this as an infinite lottery with each positive integer appearing on exactly one ticket. We would like to reflect this in a uniform distribution over the positive integers (indeed, proponents of the principle of indifference would seem to be committed to it), but if we assume countable additivity this is not possible. For if we assign probability 0 to a turn to each number being picked, then the sum of all these probabilities is again 0; yet the union of all of these events has probability 1 (since it is guaranteed that some number will be picked), and $1 \neq 0$. On the other hand, if we assign some probability $\varepsilon > 0$ to each number being picked, then the sum of these probabilities diverges to $\infty$, and $1 \neq \infty$. If we drop countable additivity, however, then we may assign 0 to each event and 1 to their union without contradiction.

Biased assignments to denumerable sets: Countable additivity allows one to assign uniform probability 1/n to each member of a denumerable partition (for example, 1/6 to each result of tossing a die). However, it requires one to assign an extremely biased distribution to a denumerable partition of events. Indeed, for any $\varepsilon > 0$, however small, there will be a finite number of events that have a combined probability of at least $1 - \varepsilon$, and thus the lion’s share of all the probability.

See [Seidenfeld 2001] for further discussion of countable additivity.

16. Probability—A Philosophical Overview

It is often thought that the only part of the axiomatization that is not merely conventional stipulation is the third axiom, in either its finite or countable form. (For example, it is tempting to say that it is purely conventional to set $P(\Omega) = 1$, rather than $P(\Omega) = 100$, say.) That is too quick. For each of the following involves a substantial mathematical assumption:

(i) Probabilities are defined by functions (rather than by one-many or many-many mappings).

(ii) These are functions of one variable (unlike primitive conditional probability functions, which are functions of two variables): there is just a single argument for a probability function.

(iii) Such a function is defined on a field (rather than a set with different closure conditions).

(iv) Probabilities are to be represented numerically (rather than qualitatively, or comparatively).

(v) Their numerical values are real numbers (rather than those of some other number system).

(vi) These values are bounded (unlike other quantities that are treated measure-theoretically, such as lengths).

(vii) Probability functions attain maximal and minimal values (thus prohibiting open or half-open ranges, such as $(0, 1)$ or $(0, 1]$).

For a discussion of rival theories that relax or replace (ii), (iii), (iv) and (vii), see [Fine 1973]. Complex-valued probabilities are proposed by Feynman and Cox (see [Muckenheim et al. 1986]); infinitesimal probabilities (of non-standard analysis) by Skyrms ([1980]) and Lewis ([1980]) among others; unbounded probabilities by Renyi (1970).

3.2 Conditional Probability

Kolmogorov also defines the conditional probability of A given B by the ratio formula:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0).$$

Thus, we may say that the probability that the toss of a fair die results in a 6 is 1/6, but the probability that it results in a 6 given that it results in an even number, is $1/2/1/2 = 1/3$. In straightforward applications in which the requisite unconditional probabilities are well-defined, and the denominator $P(B)$ is greater than 0, this formula seems to be impeccable.

But not all applications are straightforward, and in some these conditions are not met. Consider first the proviso that $P(B) > 0$. As probability textbooks repeatedly drum into their readers, probability zero events need not be impossible, and indeed they can be of real significance. It is curious, then, that some of the same textbooks glide over (1)'s proviso without missing a beat. In fact, interesting cases of conditional probabilities with probability-zero conditions are manifold. Consider an example due to Borel: a point is chosen at random from the surface of the earth (thought of as a perfect sphere); what is the probability that it lies in the Western hemisphere, given that it lies on the equator? 1/2, surely. Yet the probability of the condition is 0, since a uniform probability measure over a sphere must award probabilities to regions in proportion to their area, and the equator has area 0. The ratio formula thus cannot deliver the intuitively correct answer.
3. Independence

Independence is a fundamental concept in probability theory and is typically defined in terms of conditional probability. Two events are said to be independent if the occurrence of one does not affect the probability of the occurrence of the other. Formally, for events $A$ and $B$, independence is defined by the probability of the intersection of $A$ and $B$ being equal to the product of their individual probabilities:

$$ P(A \cap B) = P(A)P(B) $$

This is equivalent to the statement that the conditional probability of $A$ given $B$ is equal to the unconditional probability of $A$: $P(A|B) = P(A)$. Independence is a crucial property in many probabilistic models and is often assumed in various contexts to simplify calculations.

4. Interpretations of Probability

The interpretation of probability, or the philosophical question of what sorts of things are probabilistic, is a topic of considerable debate. Several prominent interpretations include frequentism, subjectivism, and objective Bayesianism. Frequentism interprets probability as the long-run frequency of an event, subjectivism views it as a measure of an individual's degree of belief, and objective Bayesianism sees it as a way to incorporate prior information into a posterior distribution. These interpretations have implications for how probabilities should be used in decision-making and statistical inference.
4.1 Classical Interpretation

The classical interpretation, historically the first, can be found in the works of Pascal, Huygens, Bernoulli, and Leibniz, and it was famously presented by Laplace ([1814]). Cardano, Galileo, and Fermat also anticipated this interpretation. Suppose that our evidence does not discriminate among the members of some set of possibilities—either because that evidence provides equal support for each of them, or because it has no bearing on them at all. Then the probability of an event is simply the fraction of the total number of possibilities in which the event occurs—this is sometimes called the principle of indifference. We may think of this as the rational subjective probability appropriate for someone in the evidential situation described. This interpretation was inspired by, and typically applied to, games of chance that by their very design create such circumstances—for example, the classical probability of a fair die landing with an even number showing up is \(\frac{1}{2}\). Probability puzzles typically take this means of calculating probabilities for granted.

Unless more is said, the interpretation yields contradictory results: you have a one-in-a-million chance of winning the lottery, but either you win or you don’t; so each of these possibilities has probability \(\frac{1}{2}\). We might look for a “privileged” partition of the possibilities, but we will not always find one. For example, in this case, the million-celled partition corresponding to each of the possible lottery outcomes seems more natural than the win/don’t win partition, if only because the former is more fine-grained. But Bertrand’s paradoxes ([Bertrand 1889]) show that particular problem may have competing, equally natural, partitions. They all turn on alternative parametrizations of a given problem that are non-linearly related to each other. The following example (adapted from [van Fraassen 1989]) nicely illustrates how Bertrand-style paradoxes work. A factory produces cubes with side-length between 0 and 1 foot; what is the probability that a randomly chosen cube has side-length between 0 and 1/2 a foot? The tempting answer is 1/2, as we imagine a process of production that is uniformly distributed over side-length. But the question could have been given an equivalent restatement: a factory produces cubes with face-area between 0 and 1 square-feet; what is the probability that a randomly chosen cube has face-area between 0 and 1/4 square-feet? Now the tempting answer is 1/4, as we imagine a process of production that is uniformly distributed over face-area. And it could have been restated equivalently again: a factory produces cubes with volume between 0 and 1 cubic feet; what is the probability that a randomly chosen cube has volume between 0 and 1/8 cubic-feet? Now the tempting answer is 1/8, as we imagine a process of production that is uniformly distributed over volume. What, then, is the probability of the event in question?

4.2 Logical Interpretation

The logical interpretation of probability, developed most extensively by Carnap ([1950]), sees probability as an extension of logic. Traditionally, logic aims to distinguish valid from invalid arguments by virtue of the syntactic form of the premises and conclusion. (E.g., any argument that has the form

\[
p \\
\text{If } p \text{ then } q \\
\text{Therefore, } q
\]

is valid in virtue of this form.) But the distinction between valid and invalid arguments is not fine enough: many invalid arguments are compelling, in the sense that the premises strongly support the conclusion—we will see an example of such an argument shortly. Carnap described this relation of “support” or “confirmation” as the logical probability that an argument’s conclusion is true, given that its premises are true. He had faith that logic, more broadly conceived, could also give it a syntactic analysis. So according to this program, probability is a measure of the degree to which a sentence supports another sentence, where this could be determined by the syntactic forms of the sentences themselves.

The program did not succeed. A central problem is that changing the language in which items of evidence and hypotheses are expressed will typically change the confirmation relations between them—for example, adding further predicates or names to a given language will typically revise how probabilities are shared around individual sentences. Moreover, Goodman ([1983]) showed that inductive logic must be sensitive to the meanings of words, for syntactically parallel inferences can differ wildly in their inductive strength. For example,

\[
\text{All observed snow is white.} \\
\text{Therefore, all snow is white.}
\]

is an inductively strong argument: its premise gives strong support to its conclusion. However,

\[
\text{All observed snow is observed.} \\
\text{Therefore, snow is observed.}
\]

is inductively weak, its premise providing minimal support for its conclusion. It is quite unclear how a notion of logical probability can respect these intuitions.

4.3 Frequency Interpretations

Frequency interpretations date back to Venn ([1876]). Gamblers, actuaries and scientists have long understood that relative frequencies bear an intimate relationship to probabilities. Frequency interpretations posit the most intimate relationship of all: identity. In a sound bite, probabilities are relative frequencies according to this view. Thus, the probability of ‘6’ on a die that lands ‘6’ 3 times out of 10 tosses is, according to the frequentist, 3/10. In general:

the probability of an outcome \(A\) in a reference class \(B\) is the proportion of occurrences of \(A\) within \(B\).

Frequentism is still the dominant interpretation among scientists who seek to capture an objective notion of probability, heedless of anyone’s beliefs. It is also the philosophical position that lies in the background of the classical Fisher/Neyman-Pearson approach that is used in most statistics textbooks. Frequentism does, however, face some major objections. For example, a coin that is tossed exactly once yields a relative frequency of heads of either 0 or 1, whatever its true bias—an instance of the infamous ‘problem of the single case’. A coin that is tossed twice can only yield relative frequencies of 0, 1/2, and 1. And in general, a finite number \(n\) of tosses can only yield relative frequencies that are multiples of \(1/n\). Yet it seems that probabilities can often fall between these values. Quantum mechanics, for example, posits irrational-valued probabilities such as \(1/\sqrt{2}\).
Some frequentists (notably Reichenbach [1949] and von Mises [1957]) address this problem by considering infinite reference classes of hypothetical occurrences. Probabilities are then defined as limiting relative frequencies in suitable infinite sequences of trials. Von Mises offers a sophisticated formulation based on the notion of a \textit{collective}: an (hypothetical, or virtual) infinite sequence of "attribution" (possible outcomes) of a specified experiment that is performed infinitely often. He goes on to lay down two requirements for such an infinite sequence \( \omega \) to be a collective. Call a \textit{place-selection} an effectively specifiable method of selecting indices of members of \( \omega \), such that the selection or not of the index \( i \) depends at most on the first \( i-1 \) attributes. The axioms are:

\textbf{Axiom of Convergence}: the \textit{limiting relative frequency of any attribute exists}.

\textbf{Axiom of Randomness}: the \textit{limiting relative frequency of each attribute in a collective \( \omega \) is the same in any infinite subsequence of \( \omega \) which is determined by a place selection}.

The probability of an attribute \( A \), relative to a collective \( \omega \), is then defined as the limiting relative frequency of \( A \) in \( \omega \).

Collectives are abstract mathematical objects that are not empirically instantiated, but that are nonetheless posited by von Mises to explain the stabilities of relative frequencies in the behaviour of actual sequences of outcomes of a repeatable random experiment. Church ([1940]) renders the notion of a place selection as a recursive function.

If there are in fact only finitely many trials of the relevant type, then this kind of frequentism requires the actual sequence of outcomes to be extended to a hypothetical or 'virtual' sequence. This creates new difficulties. For instance, there is apparently nothing that determines how the coin in my pocket would have landed if it had been tossed indefinitely—\textit{it could} yield any hypothetical limiting relative frequency that you like. Moreover, a well-known problem for any version of frequentism is the \textit{reference class problem}: relative frequencies must be relativized to a \textit{reference class}. Suppose that you are interested in the probability that Collingwood will win its next match. Which reference class should you consult? The class of all matches in Collingwood's history? Presumably not. The class of all recent Collingwood matches? That's also unsatisfactory: it is somewhat arbitrary what counts as 'recent', and some recent matches are more informative than others regarding Collingwood's prospects. The only match that resembles Collingwood's next match in every respect is that match itself. But then we are saddled again with the problem of the single case, and we have no guidance to its probability in advance. The reference class problem can also be a very practical problem—insurance companies face it on a daily basis. After all, the premiums that they set for a given individual are based on frequencies of claims of people of that type; but the individual is a member of many classes of people, whose relevant frequencies may differ wildly.

4.4 Propensity Interpretations

Propensity interpretations, like frequency interpretations, regard probability as an objective feature of the world. Probability is thought of as a physical propensity, or disposition, or tendency of a system to produce given outcomes. This view, which originated with Popper ([1959b]), was motivated by the desire to make sense of single-case probability attributions on which frequentism apparently founded, particularly those found in quantum mechanics. Propensity theories fall into two broad categories. According to \textit{single-case} propensity theories, propensities measure a system's tendencies to produce given outcomes; according to \textit{long-run} propensity theories, propensities are tendencies to produce long-run outcome frequencies over repeated trials. See [Gillies 2000] for a useful survey.

Single-case propensity attributions face the charge of being untenable. Long-run propensity attributions may be considered to be verified if the long-run statistics agree sufficiently well with those expected, and falsified otherwise; however, then the view risks collapsing into frequentism, with its attendant problems. A prevalent objection to any propensity interpretation is that it is uninformative to be told that propensities are 'propensities.' For example, what exactly is the property in virtue of which this coin has a 'propensity' of 1/2 of landing heads (when suitably tossed)? Indeed, some authors regard it as mysterious whether propensities even obey the axioms of probability in the first place. (See [Hajek 2003a].)

4.5 Subjectivist Interpretations

Subjectivist interpretations—sometimes called 'Bayesian'—pioneered by Ramsey ([1926]) and de Finetti ([1937]), see probabilities as \textit{degrees of belief} or \textit{credences} of appropriate agents. These agents cannot be actual people since, as psychologists have repeatedly shown, people typically violate probability theory in various ways, often spectacularly so. Instead, we imagine the agents to be ideally rational.

But what are credences? De Finetti identifies an agent's subjective probabilities with his or her betting behavior. For example,

\[
your \ probability \ for \ the \ coin \ landing \ heads \ is \ \frac{1}{2}
\]

if and only if

\[
you \ are \ prepared \ to \ buy \ or \ sell \ for \ 30 \ cents \ a \ ticket \ that \ pays \$1 \ if \ the \ coin \ lands \ heads, \ nothing \ otherwise.
\]

All of your other degrees of belief are analyzed similarly.

The analysis has met with many objections. Taken literally, it assumes that opinions would not exist without money, and moreover that you must value money linearly; but if it is just a metaphor, then we are owed an account of the literal truth. Even if we allow other prizes that you value linearly, problems remain. For your behavior in general, and your betting behavior in particular, is the result of your beliefs and desires working in tandem; any such proposal fails to resolve these respective components. Even an ideally rational agent may wish to misrepresent her true opinion; or she may particularly enjoy or abhor gambling; or, like a Zen master, she may lack a desire for worldly goods altogether. In each case, her betting behavior is a highly misleading guide to her true probabilities.

A more sophisticated approach, championed by Ramsey, seeks to fix an agent’s utilities and probabilities simultaneously by appeal to her preferences. Suppose that you have a preference ranking over various possible states of affairs and gambles among them, meeting certain conditions required by rationality (for example, if you prefer A to B, and B to C, then you prefer A to C). Then we can prove a ‘representation’ theorem: these preferences can be represented as resulting from an underlying probability distribution and utility function. This approach avoids some of the objections to the betting interpretation, but not all of them. Ramsey’s method essentially appeals to preferences over gambles, raising again the concern that the wrong quantities are being measured. And notice that the representation theorem does not show that rational agents’
false positives. Bayesians are likely to see it as a flaw in how probability is used. However, it is not clear that this is a flaw. In fact, it is not clear that it is a flaw at all.

In the end, it seems that the Bayesian approach to probability is more flexible and more powerful than the classical approach. It allows for a more rational and informed decision-making process. However, it is important to remember that there are many different approaches to probability, and that each has its own strengths and weaknesses.
On January 28, 1986 at 11:38 A.M., the space shuttle Challenger was launched in Florida. Seventy-three seconds later it exploded, setting back the American manned space program by years. Managers made the decision to launch, against the advice of engineers, on the basis of a superficial and flawed analysis of the probability that the two solid rocket motors would fail at low temperatures, leading to a serious underestimate of that probability (Dalal et al. 1989). Lacking a clear conception of probability—and with it, a well understood, universally accepted methodology for determining probabilities—otherwise careful engineers and managers resorted to ad hoc calculations and dubious rules of thumb that resulted in tragedy. In particular, I believe that none of the parties concerned truly understood the notion of the single-case probability of disaster that was appropriate for Challenger. And yet its launch that day, in exactly the conditions that prevailed, was by its very nature unrepeatable.

Nor have the scientists (even at Caltech!) succeeded in understanding probability. I used to set my students the following question on the final exam for my course:

In Feynman’s Lectures on Physics, Volume 1, we find the following “definition” of probability:

By the “probability” of a particular outcome of an observation we mean our estimate for the most likely fraction of a number of repeated observations that will yield that particular outcome.

There are many problems with this definition. Briefly indicate several of them.

Now that you, dear reader, have seen some of the problems with frequentism, you should be able to make a good start on this question. Here are some further hints:

- the definition is circular;
- it is easy to come up with cases in which there is more than one “most likely fraction”;
- irrational probabilities, such as $1/\sqrt{2}$ are excluded—yet our best physical theory, quantum mechanics, freely assigns such probabilities!

Finally, to bring home the subjective interpretation of probability in a way that I hope the students will never forget, I used to give them a multiple-choice test with a twist. Rather than choosing a correct answer, they had to assign **credences** to each possible answer. The test began with the following explanation:

Each of the following questions has exactly one correct answer among the choices a–d. I would like you to assign subjective probabilities to each of the choices, representing in each case your own probability that that choice is correct. For example, suppose you are nearly certain that b is the correct answer to a given question, and the other choices look about equally implausible to you. Then you might represent your opinion as follows:

a. 0.01
b. 0.97
c. 0.01
d. 0.01

---

2 David Dowse of Monash University devised a similar “probabilistic football betting” system, and I am grateful to him for suggesting the scoring rule.

---

16. Probability—A Philosophical Overview

For each question, you will receive a score determined by the probability you gave to the correct answer. Let that probability be $p$. Your score for that question will be:

$$\text{Score} = 1 + \frac{1}{2} \log_2 p$$

Thus, if you give probability 1 to the correct answer to a question, and 0 to the rest, you get a perfect score of 1 for that question; if you give 0 to the correct answer, you get a score of $-\infty$ for that question. (Totals less than 0 will be rounded up to 0.) Make sure your probabilities for a given question are nonnegative, and add up to 1—otherwise you get 0 for that question automatically!

You, dear reader, might like to try your hand at the first question on my test, reprinted below.

Good luck!

Q1. Let $\Omega$ be a non-empty set. Which of the following provides a correct characterization of a set $F$ of subsets of $\Omega$ being a sigma algebra on $\Omega$?

a. $\Omega \in F$, if $A \in F$, then $\neg A \in F$, and if $A_1, A_2, \ldots$ is a sequence of pairwise disjoint sets, each one $\in F$, then their countable union $\bigcup \mathcal{A}_n \in F$.

b. $F$ is non-empty, closed under complementation (with respect to $\Omega$) and under countable intersection.

c. $\emptyset \in F$, if $F$ is closed under complementation (with respect to $\Omega$) and under finite union.

d. $F$ is the power set of $\Omega$.

Postscript: almost every year at least one student would get a score of $-\infty$.

Acknowledgements Many thanks to Ned Hall, who years ago slaved with me on a research proposal that was an early ancestor of some parts of this paper, and also for letting me borrow from parts of our ([Halécki/Hall 2002]). Thanks also to Bonnie Gold and the editorial committee for this project for perceptive comments on earlier drafts, and to Ralph Mills for his meticulous proof-reading.

Bibliography


