Probability: a questionable science of the uncertain

LEONARD M. WAPNER

Fate laughs at probabilities.
—Edward Bulwer-Lytton, 1st Baron Lytton

Introduction

Scientific inquiry involves quantitative measurements of length, mass, and time, for which actual physical standards of comparison exist. No such standards exist when measuring *likelihood*, essential to the understanding of statistics, quantum mechanics and all forecasting sciences. Does mathematical probability convey the likelihood of actual, real-world events in a precise way?

Prior to the 17th century, chance events were generally thought to be controlled by the gods or other supernatural forces. Mathematical probability had its beginnings in the mid-seventeenth century, with the publication of Italian mathematician Girolamo Cardano's *Liber de ludo aleae* (The book on games of chance), after which the theory developed rapidly. In 1933 Russian mathematician Andrei Kolmogorov axiomatised mathematical probability, thus providing it a solid mathematical foundation.

All laws of probability are derived from Kolmogorov's three simply-stated axioms.

Axiom 1: For any event E, $P(E) \geqslant 0$.

Axiom 2: For sample space S, $P(S) = 1$.

Axiom 3: For mutually exclusive events A and B, $P(A + B) = P(A) + P(B)$.

Any function P satisfying these axioms may be called a probability function. But how and why should any such P quantify our notion of likelihood for real-world events? Probability may have full status as a mathematical discipline; but for it to qualify as a science it must reliably describe actual, not just theoretical, events. Of particular interest are hypothetical future events of world significance.

- What is the probability extraterrestrial life will be discovered this decade?
- What is the probability the Riemann hypothesis will ultimately be proved true?
- What is the probability of complete human extinction by the end of this century?

Why should the likelihood of such events, none of which have relative frequencies, be addressed by some function P, merely because P satisfies the three axioms? For such events, can likelihood be quantified? The answer is *Yes*, with stipulations.

https://doi.org/10.1017/mag.2022.117 Published online by Cambridge University Press

A personal anecdote:

> Years ago, after surgery at a local medical centre, I returned to my doctor for follow-up, at which time I asked about the risk of complications. He replied dismissively, 'Statistically it's around 5%. For you it's 100% or 0%. Don't worry about it!'
>
> I found his rendition of *Que Será, Será* to be annoyingly off-key. But it did raise questions regarding the quantification of likelihood.

A discussion of three common interpretations of probability follows. However, it is only the third—subjective probability—which properly applies to actual events. And it comes with caveats. We conclude with a discussion of highly prevalent, theoretical (and possibly actual) events for which no possibility of any interpretation exists so as to satisfy the axioms. In such cases are we obligated to concede the rules of chance back to the gods? The reader decides.

The classical interpretation

Classical probability, credited to French mathematician Pierre-Simon Laplace, interprets probability as a ratio of the number of successful outcomes to the number of all possible outcomes, assuming all possible outcomes are equiprobable. The notion relies on the *Principle of insufficient reason* (PIR), attributed to Swiss mathematician Jacob Bernoulli.

> If we have no reason to expect one outcome to an experiment over another, then all possible outcomes should be assigned the same probability.

'Insufficient reason' typically involves symmetry or lack of information. So, for rolling a standard six-sided die,

$$P(1) = P(2) = P(3) = \ldots = P(6) = \tfrac{1}{6}$$

by symmetry. If Players A and B are to play in an upcoming tennis match, and if you know nothing of their relative strengths, you might, due to lack of information, assign a probability of $\frac{1}{2}$ to Player A winning.

For the simplest of theoretical cases, classical probability works well, as long as we ignore its circular nature (probability interpreted with reference to equiprobable events). But overall, there are significant objections, many of which relating to the PIR. British economist John Maynard Keynes devotes an entire chapter to the PIR in *A treatise on probability* [1, ch. 4]. He changes the name to the *Principle of indifference* (PI), writing that the original description is 'clumsy and unsatisfactory' [1, p. 44]. He also cautions that without qualification, PI can produce contradictions and paradoxes. As one example, assume we have no information as to the colour of a certain book. It must be either red or not red and with no evidence of either, PI suggests $P(\text{red}) = P(\text{not red}) = \frac{1}{2}$. But a similar argument supports $P(\text{green}) = P(\text{not green}) = \frac{1}{2}$. Similarly $P(\text{blue}) = P(\text{not blue}) = \frac{1}{2}$.

It follows that $P(\text{red}) = P(\text{green}) = P(\text{blue}) = \frac{1}{2}$, an impossibility since the sum of probabilities associated with a finite number of mutually exclusive outcomes must be no greater than 1 to conform to the axioms.

The classical notion excludes irrational probabilities, and thus will not admit the continuous variation of likelihood. Since almost all numbers in [0, 1] are irrational, their exclusion is significant, to say the least.

Most importantly, classical probability cannot be applied to actual events where perfect symmetry is non-existent. Realistically, no actual coin or die is so perfectly shaped as to admit equiprobable outcomes. Hypothetical, real-world events such as those listed in the previous section also fall outside the scope of classical probability.

The frequentist interpretation

The frequentist (empirical) interpretation of $P(E)$, systematised by English mathematician John Venn, requires we imagine the experiment repeated infinitely many times. The probability $P(E)$ is defined as the theoretical relative frequency of occurrence of E. That is, $P(E)$ is the limiting value of the experimental relative frequency of occurrence of E as the number of trials increases indefinitely. But how does one ascertain the limiting value? Imagine a bent, almost certainly biased coin, being tossed repeatedly to determine its probability of landing heads. No finite number of tosses is guaranteed to produce the limiting relative frequency of heads. And if it did so, how would one know it?

There are additional cautions. A sequence of relative frequencies need not converge. As an example, consider $\{a_n\}$, a sequence of 1s and 0s such that

$$a_1 = 1.$$

$$\text{For } n \geqslant 2, a_n = \begin{cases} 0 & \text{if } 3^{k-1} + 1 \leqslant n \leqslant 3^k \quad \text{for some odd } k, \\ 1 & \text{if } 3^{k-1} + 1 \leqslant n \leqslant 3^k \quad \text{for some even } k. \end{cases}$$

The first 27 terms are

1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.

If k is odd, the relative frequency of 1s for the first 3^k terms is less than or equal to

$$\frac{3^{k-1}}{3^k} = \frac{1}{3}.$$

But if k is even, the relative frequency of 1s for the first 3^k terms is greater than or equal to

$$\frac{3^k - 3^{k-1}}{3^k} = \frac{2}{3}.$$

It follows that the relative frequency of 1s fails to converge. This sequence, unlikely to occur randomly, is no less likely to occur than any other specified sequence.

In some cases, the convergence of a relative frequency may depend on how the outcomes are ordered. In *Fifteen arguments against hypothetical frequentism* [2], Alan Hájek, professor of philosophy at the Australian National University, provides an example from which the following is adapted.

A man walks along a north-south path, reversing direction from time to time. Each time he changes direction he tosses a coin, recording the outcome with respect to both time and displacement north. Figure 1 shows the outcomes as they occur in time (horizontal axis) and space (vertical axis).

What can we say about the relative frequency of heads? Scanning from left to right with respect to time, the sequence of toss outcomes follows the pattern HTHTHT …, indicating heads occurs with a relative frequency of $\frac{1}{2}$. But scanning vertically with respect to north displacement the pattern is HHTHHT …, giving heads a relative frequency of $\frac{2}{3}$. So, is the relative frequency of heads $\frac{1}{2}$ or $\frac{2}{3}$? There simply is no reason why time (or space) should be the preferred ordering for the sequence. The limiting relative frequency will depend on the order of the toss outcomes while the likelihood of heads (or tails) should not.

Theoretical experiments, such as the random toss of a perfectly fair coin, are replicable and thus possess theoretical relative frequencies. But actual experiments are not because the conditions under which they are performed cannot be exactly duplicated. For such experiments the frequentist notion cannot be applied. And once again hypothetical, real-world events are beyond the reach of empirically defined probabilities. *What is the probability of a universal cancer cure being discovered this decade*? Frequentists do not assign probabilities to such events.

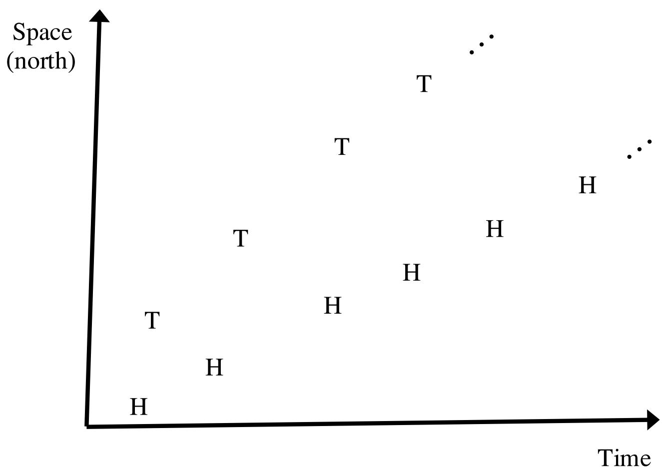

FIGURE 1: Two possible relative frequencies

https://doi.org/10.1017/mag.2022.117 Published online by Cambridge University Press

The subjective interpretation

Subjective probability is a strength of belief, derived from one's personal judgement of the likelihood of an event. It is an opinion based on past experience and information at hand. Italian probabilist and statistician Bruno de Finetti boldly promotes subjective probability in writing:

> … we follow and sustain here only subjective probabilities exist – that is, the degree of belief in the occurrence of an event attributed by a given person at a given instant and with a given set of information. This is in contrast to other conceptions that limit themselves to special types of cases in which they attribute meaning to 'objective probabilities' (for instance, cases of symmetry as for dice etc., 'statistical' cases of 'repeatable' events, etc.) [3, p. 3].

So the subjective interpretation requires human contemplation of the event; otherwise there is no probability, in which case the event either occurs or it doesn't.

Today's scientific research relies heavily on inferential statistics, of which there are two philosophies—frequentist inference (based on the frequentist interpretation of probability) and Bayesian inference, where subjective probabilities are assigned to hypothetical, non-repeatable events. Being a foundation of Bayesian theory, subjective probability plays a significant role in statistical research.

And how is subjective probability to be measured? The maxim, 'Something is only worth what someone is willing to pay for it' suggests subjective probability of actual, one-time events can be measured by how much one is willing to wager that the event will occur. The best example is that of pari-mutuel wagering. At a racetrack, bets placed on a horse to win a race are converted to and posted as odds, reflecting the crowd wisdom of the likelihood of the horse winning. So, if the current *fractional odds* on Nibbles to win are 5-2 (bettors risk losing 2 units to win 5), then the bettors' subjective probability of Nibbles winning (ignoring the bookmaker's margin) is $\frac{2}{2+5} = \frac{2}{7}$.

(In *decimal odds* format, Nibbles' odds would show as 3.50, representing the total potential return (including the stake) on a 1 unit bet. Note the relationship between decimal odds and subjective probability: $\frac{1}{3.50} = \frac{2}{7}$.)

Modern technology provides *prediction markets* as virtual online malls for participants to bet on future outcomes (election results, market behaviour, weather events, etc.) materialising. The value of the bet reflects the subjective probability (crowd wisdom) of the event actually occurring. Assume the market posts a question such as, 'Will Scotland become an independent country by the end of this decade?' A bidder may be willing to pay at least .60 units now to purchase the 'Yes' share of a contract, securing a 1 unit payoff should this event occur. This equates to the bidder assigning a 60% subjective probability to Scottish independence. If another bidder is willing to pay at least .40 units (40% subjective probability) to purchase the 'No' share, then a contract worth 1 unit is created by these two bidders. Additional contracts are formed whenever the bid for a 'Yes' share and the bid for a 'No' share sum to 1 unit. Shares may also be traded, sold and bought at the current market value which reflects the current likelihood of the event materialising. Ultimately, the 1 unit payoff associated with each contract goes to either the first bidder or the second, depending on the actual outcome.

Compared to the classical and frequentist notions, subjective probability is the only one applicable to one-time actual (as opposed to theoretical) events. Nevertheless, subjective probability is problematic. It fails to exist without human contemplation. Physical and biological events in our universe will continue to occur with some degree of likelihood even if human existence ends, at which point subjective probability ceases to be meaningful.

As a science, subjective probability is as descriptive (perhaps more so) of the belief holder as it is of the event in question, depending in large part on an individual's intelligence and rationality. It therefore fails to be a well-defined characteristic of the event in question. And the axioms might be violated because individuals assigning subjective probabilities need not be aware of nor be compelled to respect the axioms. Such assignments might be considered irrational and the probabilities so assigned to be inadmissible.

It would seem that at least one of the three interpretations of probability discussed here (classical, frequentist, subjective) would apply to most, if not all events, both theoretical and actual. Remarkably, there exist events (theoretical and potentially actual) for which no probability *by any interpretation* exists so as to satisfy the axioms. Examples of such are closely related to non-measurable sets—sets for which no Lebesgue measure can be assigned. The example to follow is highly contrived; yet it suggests similar events, actual and theoretical, may be highly prevalent.

The Vitali event

Just over a century ago, Italian mathematician Giuseppe Vitali defined a set of numbers in the interval [0, 1], later named the *Vitali* set, to which no Lebesgue measure can be assigned. The Vitali set can also be defined on a circle C of unit circumference. To this set there exists a related event for which no probability exists to satisfy the Kolmogorov axioms.

Let A and B be any two points on C, separated by an arc of length s. We call A and B equivalent ($A \sim B$) if s is a rational number. In other words $A \sim B$ if, and only if, $s \in Q$, where Q denotes the set of rational numbers. This effectively partitions the circle C into uncountably many equivalence classes, each containing a countable number of equivalent points. As allowed by the Axiom of Choice, we may select exactly one point from each equivalence class. The set of all such points forms the Vitali set V, a point set on circle C.

Now let Q^* consists of all rational numbers in [0, 1). Since the number of points in Q^* is countable, we can describe it as

$$Q^* = Q \cap [0, 1) = \{q_1, q_2, q_3, \dots\}.$$

Next, for $i = 1, 2, 3, \dots$, let V_i denote the point set obtained by rotating V clockwise through an arc of length $q_i \in Q^*$. The infinitely countable collection of V_i, all congruent to V, forms a partition of the circle. That is,

$$\bigcup_{i=1}^{\infty} V_i = C \text{ and } V_i \cap V_j = \varnothing \text{ for } i \neq j.$$

Now flick the spinner shown in Figure 2.

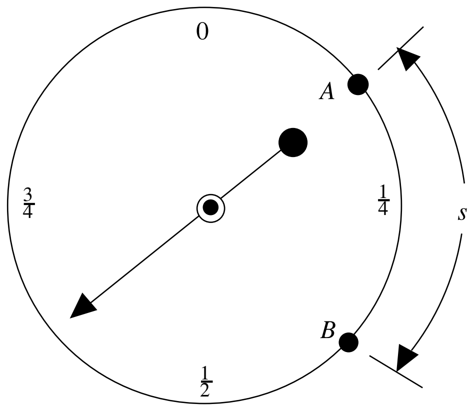

FIGURE 2: Vitali spinner

There's a chance the spinner will randomly land on a point in V. But with what probability? What is the value of $P(V)$? Because each V_i is congruent to V and the spinner lands uniformly on C,

$$\sum_{i=1}^{\infty} P(V_i) = \sum_{i=1}^{\infty} P(V).$$

If $P(V) = 0$ the sum is 0. If $P(V) > 0$ the sum is infinitely large. Either way, Axiom 2, requiring the sum be 1, is violated. Under the Kolmogorov axiomatization, $P(E)$ cannot possibly exist.

The Vitali set V and associated Vitali event are admittedly contrived, raising a question as to the prevalence of such events. We are compelled to ask how the number of non-measurable sets compares to the number of measurable ones. To investigate, we'll consider both types of sets in the interval [0, 1].

As customary, let c, the *cardinality of the continuum*, denote the cardinality of the set of all points in [0, 1]. The Cantor set, a subset of [0, 1], is of cardinality c, yet is easily shown to be of measure zero. There are 2^c

https://doi.org/10.1017/mag.2022.117 Published online by Cambridge University Press

subsets of the Cantor set, all of which are measurable since a subset of a set of measure zero must itself be measurable. Because the Cantor set is a subset of [0, 1] we infer there are at least 2^c measurable subsets in [0, 1] which can be no more in number than 2^c, the total number of subsets of [0, 1]. Therefore the number of measurable subsets of [0, 1] is 2^c.

Interestingly, the Cantor set can also be used to determine the number of non-measurable subsets of [0, 1]. Let A denote any non-measurable set in [1, 2]. The union of any such set with a subset of the Cantor set is non-measurable because a non-measurable set combined with one of measure zero remains non-measurable. Since there are 2^c subsets of the Cantor set, there are 2^c such unions, all distinct. At this point we have 2^c non-measurable subsets of [0, 2]. Dividing by two (shrinking by a factor of two) takes us to [0, 1] where we now have non-measurable subsets of [0, 1].

Non-measurable point sets being as prevalent as measurable point sets suggests that actual events to which no probability can be assigned may be far more prevalent than expected.

Closing thoughts

- Actual events, if precisely specified, are one-off and have no relative frequency of occurrence.
- Subjective probability is more a strength of belief than a characteristic of the event being considered.
- There may be actual events for which no probability exists satisfying the axioms. The prevalence of such events may be far greater than one might expect. This raises questions regarding the necessity of and our reliance on the axioms for actual events.

Nineteenth century German physiologist Emil du Bois-Reymond, in a speech to the 1872 Congress of German Scientists and Physicians, described scientific areas of inquiry which transcend our ability to comprehend them [4]. He popularised the Latin phrase *ignoramus et ignorabimus* ('we do not know and will not know'). It was this pessimistic tone which later prompted German mathematician David Hilbert to issue his now famous challenge, *Wir müssen wissen – wir werden wissen* ('We must know – we will know'). Du Bois-Reymond singled out the ultimate nature of matter, the origin of motion and the origin of sensations as being the most difficult to describe. Perhaps he should have included the likelihood of chance events on his list.

While giving a lecture, Bertrand Russell once remarked, 'probability is the most important concept in modern science, especially as nobody has the slightest notion of what it means' [5, p. 587]. Perhaps he overstates it. But he reminds us that probability is far from exact; it has no single interpretation and there are no standards of measurement. Fortunately we need not concede the rules of chance back to the gods. The aphorism 'All models are wrong but some are useful' certainly applies to mathematical probability. For now, it's *probably* the best we can do.

Acknowledgement

The author gratefully acknowledges the anonymous referee for constructive suggestions.

References

1. J. M. Keynes, *A treatise on probability*, Macmillan (1921).
2. A. Hájek, Fifteen arguments against hypothetical frequentism, *Erkenntnis* **70**(2) (2009) pp. 211-235.
3. B. de Finetti, *Theory of probability: A critical introductory treatment*, Wiley (2017).
4. E. du Bois-Reymond, *Über die Grenzen des Naturekennens* (The limits of science), Wentworth (2018).
5. E. T. Bell, *The development of mathematics*, McGraw-Hill (1945).

10.1017/mag.2022.117 © The Authors, 2022
Published by Cambridge University Press
on behalf of The Mathematical Association

LEONARD M. WAPNER
Division of Mathematical Sciences, El Camino College, Torrance, California 90506 USA
e-mail: *lwapner@elcamino.edu*

In the Pipeline for March 2023

The Mathematikum in Giessen	*Martin Buhmann*
Sums of the first odd integers	*Chris Sangwin*
Hardy's inequality for averages	*G. J. O. Jameson*
Happy functions, digit maps, diophantine equations, and proofs of some mathematical memes	*Niphawan Phoopha, Prapanpong Pongsriiam, Phakinkon Phunphayap*
Desargues' involution theorem: from history to applications	*Alexandru Gîrban*
Visualizing connections between types of polygonal number	*Umit Sarp*
Cryptography based on algebraic perpendicularities	*Timo Tossavainen*
Wolstenholme's inequality and its relation to Barrow's and Garfunkel-Bankoff inequalities	*Martin Lukarevski*
More characterisations of parallelograms	*Mowaffaq Hajja, Panagiotis T. Krasopoulos*
Multiplication is to addition as addition is to what?	*Howard Sporn*
Integrating sine and cosine Maclaurin remainders	*Russell A. Gordon*
Cauchy and his modern rivals	*William Dunham*

10.1017/mag.2022.118 © The Authors, 2022