Philosophia Mathematica (III) **19** (2011), 308–322.
doi:10.1093/phimat/nkr022 Advance Access publication August 17, 2011

Objective Probabilities in Number Theory[†]

JORDAN ELLENBERG[*] AND ELLIOTT SOBER[**]

Philosophers have explored objective interpretations of probability mainly by considering empirical probability statements. Because of this focus, it is widely believed that the logical interpretation and the actual-frequency interpretation are unsatisfactory and the hypothetical-frequency interpretation is not much better. Probabilistic assertions in pure mathematics present a new challenge. Mathematicians prove theorems in number theory that assign probabilities. The most natural interpretation of these probabilities is that they describe actual frequencies in finite sets and limits of actual frequencies in infinite sets. This interpretation vindicates part of what the logical interpretation of probability aimed to establish.

1. Preliminaries

Very few philosophers nowadays have much time for the logical interpretation and the actual-frequency interpretations of probability. And the hypothetical-frequency interpretation is almost as unpopular. We begin with a brief and opinionated review of why.

The logical interpretation was inspired by an analogy with formal logic — namely, that probability represents a weakened form of deductive entailment. As the value of $\Pr(B|A)$ approaches unity, the relationship of A and B is supposed to resemble more and more the relationship that obtains when A logically entails B. Whereas deductive entailment is a yes-or-no affair, probability is said to represent the degree to which one proposition entails another. When A entails B, A provides the strongest assurance that B is true; when A partially entails B, the guarantee is weaker. According to the logical interpretation, statements that assign probabilities to propositions are supposed to express logical truths; both 'A entails B' and 'A partially entails B' are supposed to be analytic.

[†] This work was partially funded by support to ES from the William F. Vilas Trust, University of Wisconsin, Madison.

We are grateful to Hayley Clatterbuck, Kenny Easwaran, Branden Fitelson, Mike Titelbaum, Mark Sapir, and other commentors at MathOverflow.net, and to the referees of this journal for useful discussion.

[*] Mathematics Department, University of Wisconsin, Madison, Wisconsin 53706, U.S.A. ellenber@math.wisc.edu

[**] Philosophy Department, University of Wisconsin, Madison, Wisconsin 53706, U.S.A. ersober@wisc.edu

Carnap [1950b] held that there are two concepts of probability — objective frequencies and rational degrees of belief; he developed the logical interpretation as an explication of the latter. Carnap's approach was syntactic; within his framework, a given proposition can have different probabilities depending on the formal language in which it is expressed. Is this feature of Carnap's account enough to sink it? We think not. The parallel with deductive logic is instructive. Logical implication is a semantic relation, but it still is possible to describe a syntactic relation that holds between sentences in a formal language that mirrors this semantic relation. There is nothing wrong with syntactic treatments of logical implication in a language, so long as we do not forget that they are specific to this or that language. One reason that syntactic treatments of logical implication have their place is that they obey an important constraint. If a language is enriched by adding new constants or predicates or logical operators, this may have the effect that the new sentences thus created are related by logical implication, but the enrichment should never nullify the relations of logical implication that obtained between sentences already present in the weaker language.[1] For example, if A entails B in sentential logic, A must also entail B in first-order quantifier logic. Carnap's syntactic program for probability violated this constraint. The concept of a family of predicates is central to Carnap's approach. The predicates in a family must be exclusive and exhaustive. It turns out that different values attach to Pr(individual a is green), depending on whether you use a language in which 'green' is one of n color predicates, or use another language that is otherwise the same except that 'light green' and 'dark green' are two of the $n + 1$ color predicates. Other problems arise as well. Here we will mention just one of them, which pertains to the question of whether Carnap succeeded in showing that probability is a *logical* notion. Carnap appeals to symmetry considerations to justify assigning equal prior probabilities to sentences that differ only via permutations of the names of individuals, but these considerations cannot be said to embody *logical* truths [Hájek, 2009].

Even if these specific problems had not reared their ugly heads, Carnap's syntactic approach would still be open to the charge that it fails to get at what is fundamental about probability. Imagine a set of syntactic characterizations of logical implication, each correct for a particular formal language, that is not anchored to and unified by a semantic characterization of the logical implication relation. Something fundamental would be missing. Now consider the proposition that an atom of Uranium 238 will

[1] Kenny Eswaran (personal communication) raises the question of whether enriching a language can create new relations of logical implication among sentences in the old language. He points out that Pierce's Law — that $((A \rightarrow B) \rightarrow A) \rightarrow A$ — is not provable in the fragment of classical logic containing just the conditional, but is provable with the double-negation rule.

decay in the next 10^{10} years. Of course, if you have a rational degree of belief p in this proposition, you may be able to craft a language in which that probability value is somehow reflected in syntactic features of the language you construct. But, as Socrates said to Euthyphro, do not put the cart before the horse. It is not the language you speak that makes your degree of belief rational. Probability has nothing essentially to do with language.[2]

Just as the logical interpretation of probability has fallen into disrepute, the same is true of various objective interpretations that attempt to capture what is going on when probability assignments are empirical. The actual-frequency interpretation of probability [Venn, 1876] says that the conditional probability $\Pr(B|A)$ is the actual frequency with which B is true when A is true. It runs into trouble when we toss a fair coin an odd number of times. In this case one would like to say that

$$\Pr(\text{the coin lands heads} \mid \text{the coin is tossed}) = 1/2, \tag{1}$$

but if the coin is tossed an odd number of times, one *cannot* obtain 50% heads; and even if the fair coin is tossed an even number of times, one *might* not obtain that result. The problem is not restricted to gambling devices. Consider, for example, the idea from Mendelian genetics that

$$\Pr(\text{offspring is a heterozygote at locus } L \mid \text{parents are both heterozygotes at locus } L) = 1/2.$$

This does not require that half the offspring of such a parental pair are heterozygotes.

There is another frequency interpretation of probability that may seem to do better. Although a fair coin can be tossed an odd number of times, maybe what is true of a fair coin is that its frequency of heads would converge on 1/2 if it were tossed repeatedly. This is the hypothetical-frequency interpretation of probability. It equates the probability's being 1/2 with:

$$\text{Freq}(\text{the coin lands heads} \mid \text{the coin is tossed } n \text{ times}) \text{ approaches } 1/2 \text{ as } n \text{ approaches infinity.} \tag{2}$$

The problem is that a fair coin can land heads every time it is tossed [Skyrms, 1980]. Statement (1) is not equivalent with statement (2). This

[2] There are other philosophical issues that Carnap addressed by formulating theses about language, when, in fact, his basic philosophical idea has no essential connection with language. For example, [Carnap, 1950a] draws an epistemological distinction between 'internal' and 'external' questions; the latter cannot be answered by giving empirical evidence or by citing a proof; rather, external questions can be answered only by citing the fact that some answers are more useful than others. Carnap also says that answering an external question involves adopting a linguistic framework. However, there is no need to give this conventionalist epistemology a linguistic formulation [Sober, 2000].

is not to deny that (1) is equivalent with

$$\text{For any } \epsilon > 0,\ \Pr(\text{the percentage of heads in } n \text{ tosses is within } \epsilon \text{ of } 50\% \mid \text{the coin is tossed } n \text{ times}) \text{ approaches 1 as } n \text{ approaches infinity.} \quad (3)$$

However, (3) is not an *interpretation* of the probability concept used in statement (1), since (3) uses the very concept ('probability') that we are asked to clarify. An *interpretation* of probability should elucidate that concept in terms of *other* concepts that are already understood.

What interpretive options remain? Well, there are subjective interpretations of probability, according to which probability represents rational degree of belief (certainty). However, this will not appeal if you are looking for an objective interpretation. And there are propensity interpretations, according to which the coin's probability of landing heads is a propensity the coin has. Whether this is a genuine interpretation is controversial. What does 'propensity' mean? If the suggestion is that '$\Pr(B|A)$' describes the causal power of A to bring about B, then this interpretation fails to capture a host of probability statements — for example, those in which the probability of a cause conditional on an effect is described (Salmon [1984, p. 205] attributes this criticism to Paul Humphreys).[3] And then there is the 'no-theory theory', according to which objective probabilities are theoretical quantities that obey the axioms of Kolmogorov [1950] and that cannot be defined in terms of observables [Sober, 2010]. We shall not explore these options here, since our main interest is in the failed interpretations described above. We argue in what follows that the use of probabilities in number theory is good news for frequency interpretations. And it breathes new life into part of what the logical interpretation asserts.

2. Finite Sets

It is unproblematic that mathematical properties have various frequencies of occurrence in different finite domains of mathematical objects. For example, consider the statement

$$\Pr(i \text{ is prime} \mid i \text{ is a member of the set } \{3, 4, 9, 12\}) = 1/4. \quad (4)$$

We know that actual frequencies satisfy the Kolmogorov axioms. Proposition (4) comes out true if we use the actual-frequency interpretation of probability.

[3] And even when A is a causal promoter of B, it is unsatisfactory to represent the strength of this causal promotion by $\Pr(B|A)$. It would be better to calibrate causal strength in terms of the difference that A makes in B, namely as $\Pr(B|A) - \Pr(B|\neg A)$, though even this is unsatisfactory; just as cause and correlation are different, the strength of a cause is not well measured by the strength of the correlation.

If proposition (4) expresses a truth about actual frequencies, it cannot be understood as having the variable 'i' bound to a universal quantifier. It is not true that each member of the set {3, 4, 9, 12} is prime with frequency 1/4. In this respect, proposition (4) differs from proposition (1). The usual model of tossing a fair coin says that tosses are independent and identically distributed (*i.i.d.*) — *each toss* has a probability of 1/2 of landing heads. Interpreted in terms of actual frequencies, proposition (4) says that the property of being prime and the property of being a member of the set {3, 4, 9, 12} are related in a certain way; there is no universal quantifier.

3. Infinite Sets

Unlike the trivial probability statement just discussed concerning the set {3, 4, 9, 12}, there are various nontrivial theorems that number theorists have proved that assign probabilities (see, *e.g.*, [Hardy and Wright, 2008] and [Lang, 1994]). Here are two examples:

The probability that an integer has no perfect squares other than 1 among its divisors is $6/\pi^2$. (5)

The probability that a prime is congruent to 1 mod 4 is 1/2. (6)

Both these statements are understood by number theorists to refer to limits on finite sets that are made increasingly large. Statements (5) and (6) get spelled out as follows:

Pr(i has no perfect squares other than 1 among its divisors | i is one of the first n integers) approaches $6/\pi^2$ as n approaches infinity. (5*)

Pr(i is congruent to 1 mod 4 | i is one of the first n primes) approaches 1/2 as n approaches infinity. (6*)

Since statement (4) can be interpreted by using the actual-frequency interpretation of probability, should (5*) and (6*) be understood in terms of the hypothetical-frequency interpretation? In the previous section, we described a problem that the hypothetical-frequency interpretation encounters when it is applied to fair coins. No such problem arises in number theory. A fair coin can have various frequencies of heads when it is tossed n times. But the frequency of squarefrees in the first n integers has no such wiggle room. And the limits of those frequencies are perfectly determinate as well. Propositions (5) and (6) *can* be interpreted by thinking of probability as a limit of actual frequencies. And whereas real coins are never tossed infinitely many times, there actually are infinitely many integers. With respect to the actual-frequency interpretation of probability, numbers are kinder than coins.

But there are complications. A statement like (5) has no determinate truth value until it is made precise by an interpretation like (5*). A different interpretation can lead to a different conclusion. For example, one might define a function $f(n)$ by the rule

$$f(n) = n \text{ when } n \text{ is a multiple of 4; } 10n \text{ otherwise}$$

and interpret (5) as

$$\Pr(i \text{ has no perfect squares other than 1 among its divisors} \mid i \text{ is one of the integers } k \text{ satisfying } f(k) < n) \text{ approaches } 6/\pi^2 \text{ as } n \text{ approaches infinity.} \quad (5^{**})$$

The assertion (5**), just like (5*), interprets the probability in (5) as a limit of actual frequencies in finite sets that grow larger and larger, where the union of these sets is the entire set of positive integers. But (5**) is false. If (5) is true, then (5**) cannot be the right account of what (5) means. But what does (5) really mean? One might argue that (5*) is a more 'natural' interpretation of (5) than (5**) is, but arguments of this kind are difficult to make objective.

Here is an even simpler example of the same phenomenon. Consider the proposition:

The probability that a positive integer is even is 1/2. (7)

This seems uncontroversial, and seems to be backed up by the following statement about limits:

$$\Pr(i \text{ is even} \mid i \text{ is one of the first } n \text{ integers}) \text{ approaches } 1/2 \text{ as } n \text{ goes to infinity.} \quad (7^{*})$$

Proposition (7*) is the right way to spell out what (7) means, under the customary understanding of 'first'. But what if we enumerate the integers in a different order? For example, the sequence

$$1, 3, 2, 5, 7, 4, 9, 11, 6, 13, 15, 8, \ldots$$

renders (7*) false. Instead we have

$$\Pr(i \text{ is even} \mid i \text{ is one of the first } n \text{ integers}) \text{ approaches } 1/3 \text{ as } n \text{ goes to infinity.} \quad (7^{**})$$

If the probability that a positive integer is even cannot depend on how the integers are ordered, then this result is a problem. It is not considered a problem by number theorists, largely because the usual ordering of positive integers is *canonical*. It is the only ordering compatible with addition, which is to say that it is the only one such that $a + b > a$ for all positive integers b. To choose any other ordering would be willfully perverse.

Our suggestion is that (7) is an incomplete statement. Number theorists usually interpret it by thinking about (7*), which is why they judge that (7) is true. But this is not because (7**) and its ilk are false; (7**) is true (when it is understood in terms of the nonstandard sequence of integers just described), but this does not show that (7), under its intended interpretation, is false. Taking limits depends on an ordering relation, which goes unspecified in (7); (7*) is true relative to one ordering of integers, while (7**) is true relative to another. The thesis that probability statements like (5), (6), and (7) should be understood in terms of limits of actual frequencies in ever-increasing finite sets needs to recognize that such limits are relative to an ordering relation.

Another delicate point arises when we ask what the probability is that a positive integer has first digit 1. Following (5*) and (6*), one might try to compute the limit of the quantity

$$\Pr(i \text{ has first digit } 1 \mid i \text{ is one of the first } n \text{ integers}) \qquad (8)$$

as n goes to infinity, but this limit does not exist. This might lead one to assert that (8) does not have a well-defined value. On the other hand, in naturally occurring data, the frequency of numbers with first digit 1 is often well-approximated by $\log 2/\log 10$, a phenomenon known as *Benford's Law*. According to this 'law', the first digit should be 1 a little less than 1/3 of the time and larger digits should occur as leading digits with lower and lower frequency, with 9 occurring in first position less than 5% of the time [Hill, 1998]. Surprisingly, this pattern applies to a wide range of data sets, including street addresses, stock prices, population numbers, death rates, and lengths of rivers. Even certain tables of mathematical and physical constants, which presumably are not products of chance processes, obey Benford's Law.

Does Benford's Law show that it is a mistake to insist that (8) be interpreted as a limit of frequencies? No; it merely shows that in this case the apparently most natural limiting procedure does not produce an answer — an idea that we place in a wider context in the next section.

4. Beyond Simple Frequency: The Dirichlet Density

The frequencies on the finite interval $[1 \ldots n]$ considered in connection with propositions (5), (6), and (7) can also be considered as weighted frequencies on the set of all integers, in which the first n positive integers are assigned equal weights, and larger integers are assigned weight 0. But no principle commits us to this particular choice. Indeed, it is not the most popular choice among number theorists. A more common interpretation of probability is the *Dirichlet* (or *logarithmic*) density. Under this interpretation, the statement 'a random integer is contained in the set A with probability p' means that

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{\substack{a\in A\\ a<n}} \frac{1}{a} = p.$$

If the limit does not exist, we do not assign a probability to the statement 'A randomly chosen integer lies in A'.

The exact form of the Dirichlet density is not important for this paper. Just take from it the fact that, under this interpretation, probability is again a limit of frequencies (albeit weighted frequencies) on finite sets. It turns out that if a proposition has a probability under the interpretation of the previous section (usually called the *asymptotic* or *natural density*), it has a Dirichlet density too, and the values agree. But there are some propositions that have a probability under the Dirichlet interpretation but not under the naïve one: for instance, the probability that an integer has first digit 1 in the Dirichlet interpretation is $\log 2/\log 10$, just as one finds in many sets of naturally occurring numbers. So Benford's Law is not outside the limit-of-frequencies interpretation of probability.

The Dirichlet density does not require that the integers arose by a random process. In this respect it should not be surprising that Benford's Law applies both to numbers that can be interpreted as the result of chance processes (*e.g.*, death rates) and those that cannot (*e.g.*, mathematical constants). It turns out that the frequency distribution postulated by Benford's Law is the only distribution compatible with the 'scale invariance' requirement that the distribution be the same regardless of the choice of units (*e.g.*, inches, kilometers, or cubits). Not every empirical distribution has this property. For example, consider the fact that most adult human beings have heights between 50 and 79 inches; the high frequencies of 5s, 6s, and 7s as initial digits in this distribution lapses if centimeters are used instead.

If natural densities can be supplemented by Dirichlet densities, can Dirichlet densities be supplemented in turn? We take no stand on this mathematical question, though we note that mathematical practice currently does not take this further step. Dirichlet densities agree with natural densities when the natural density exists. The natural density can therefore be seen as a base case. If the natural density specifies a value for a probability, other admissible densities must agree on that value. But when the natural density is silent, others may speak up, though an argument is needed to justify what they say.[4]

5. The Interpretation of Conditional Probability

The so-called Kolmogorov definition of conditional probability states that

$$\Pr(B|A) =_{def} \Pr(A \,\&\, B)/\Pr(A). \tag{K}$$

It entails that the conditional probability is not defined when $\Pr(A) = 0$. Popper [1959], Hájek [2003], Sober [2008], and others have criticized this definition by describing simple empirical cases in which conditional probabilities make sense and have obvious values even when the conditioning proposition has a probability of zero. For example, Pr(the coin lands heads | the coin is tossed) can equal 1/2 even if you make it impossible for anyone to toss the coin (perhaps by locking it in a safe that cannot be opened). A similar objection to the Kolmogorov definition arises in number theory. Consider proposition (6) and restrict the universe to the integers. Pr(an integer is prime) = 0 which means that Pr(integer is prime and is 1 mod 4) = 0 as well, but the conditional probability is not *undefined*; it is 1/2. Instead of regarding (K) as a *definition* of conditional probability, perhaps it is better to regard it as a *statement* about conditional probability that is true when (but not only when) $\Pr(A) > 0$.

Once the sense of (K) as a definition has been discarded, it might be preferable to write the statement in the form

$$\Pr(B|A)\Pr(A) = \Pr(A \,\&\, B),$$

in which case the proposition is true regardless of whether $\Pr(A) > 0$. Interpreted in this way, the statement does not say that $\Pr(B|A)$ is always undefined when $\Pr(A) = 0$; indeed, in that case the statement is consistent with any value of $\Pr(B|A)$. This is not to say that $\Pr(B|A)$ cannot be defined objectively when $\Pr(A) = 0$. For example, consider the following propositions about positive integers

$$A : n \text{ is a positive integer at most 10.} \quad B : n \text{ is even.}$$

Surely one wants to say that $\Pr(B|A) = 1/2$, even though $\Pr(A) = 0$ under any reasonable version of 'limiting actual frequencies'. More bluntly,

[4] Although the natural density dictates to Dirichlet densities, but not conversely, the fact remains that values for natural densities are sometimes more easily derived by thinking about Dirichlet densities. For instance, we do not know of a proof of (6*) that determines the natural density directly; instead, the proofs of (6*) and its generalizations (due to Dirichlet) determine the Dirichlet density, and draw a conclusion about natural density thereby. So Dirichlet densities sometimes render natural densities epistemically accessible, even if it is the natural density that is in some ways more conceptually fundamental.

$\Pr(A|A)$ should be taken to be 1 even if $\Pr(A) = 0$. We might make an exception for cases where A is not merely an event of probability 0, but is logically impossible; it is not obvious to us that 'Pr(5 is even | 5 is even)' can be sensibly defined.[5]

6. The Logical Interpretation of Probability

As noted above, Carnap developed his logical concept of probability by discussing syntactic features of languages. But, as Carnap realized, there is more to logic than syntax. What would a logical interpretation be like if it were nonsyntactic? One possibility is the thesis that probability statements are logical truths if they are true at all. This thesis fails for a large number of empirical probability statements. It is not a logical truth that this coin is fair, or that that organism has a given probability of being a heterozygote, or that this atom has a given half-life. Even so, *some* probability statements are mathematical truths. These include theorems in number theory like (5) and (6), which, as mentioned, need to be relativized to an ordering relation. Whether these are *logical* truths depends on what one means by logic. Narrower construals of logic will have one result, broader construals (*e.g.*, that set theory is part of logic) another. We have no stake in this question about logicism.

Probability statements that are *a priori* also show up in empirical sciences. An example is the statement that

$$\Pr_M(\text{the offspring is a heterozygote at locus } L \mid \text{both parents are heterozygotes at } L) = 1/2,$$

when M is the usual Mendelian model (for other examples, see [Sober, to appear]). Of course there is more to Mendelian genetics than *a priori* statements such as this one; there is, additionally, the empirical claim that Mendelism furnishes an empirically adequate account of heredity [Fitelson, 2007].

It should be noted, however, that the fact that various probability statements are logical or mathematical truths does not provide an *interpretation* of probability; 'snipes are snipes' is a logical truth, but that does not tell you what 'snipe' means. Here the view may be supplemented with the frequency interpretation we have described. When probability statements about gambling devices, Mendelian inheritance, and atomic half-lives are

[5] Be warned, however, that some 'obviously correct' choices for defining $\Pr(B|A)$ when $\Pr(A) = 0$ lead to contradictions, as demonstrated by Easwaran [2008b]. Popper and Hájek each propose theories of how conditional probabilities get their values without recourse to Kolmogorov's 'definition'. In fact, Kolmogorov [1950] himself provides a theory of this sort as well. Easwaran argues that Kolmogorov's theory is superior to both Popper's and Hájek's.

considered, the logical and the frequency interpretations are rivals. In the context of number theory, they are allies. Probability statements in number theory describe frequencies, *and* they are *a priori*.

7. Are Number-Theoretic Probabilities Really Probabilities?

A skeptical reader might here suggest another reason that the probabilities in number theory are more amenable to frequency interpretations than are coin flips — namely, that number-theoretic 'probabilities' are not probabilities at all. A prime number is either 1 mod 4 or it is not. No chance process is involved. Our reply is that we agree that each number is necessarily 1 mod 4, or necessarily not. But it does not follow that there are no probabilities in number theory other than 0 and 1. Propositions (5), (6), and (7) are, we agree, necessarily true, but that hardly shows that they do not use the concept of probability. It is important to distinguish probabilities *in* theorems from probabilities *of* theorems; even if the latter need to be interpreted subjectively, it doesn't follow that the former do too.[6] As to the suggestion that objective probabilities apply only to the possible outcomes of chance processes, our answer is that assertion is no substitute for argument. Frequency interpretations are sufficiently ensconced in the history of proposed interpretations of probability that it would be churlish to deny that frequencies are probabilities when this interpretation actually happens to work.

A second kind of skeptic would insist that probability is by definition a measure of subjective degree of rational belief and that when number theorists say the word 'probability' but mean 'actual frequency or limit of same', their idiosyncratic usage should not be considered relevant to philosophical questions about probability. To this subjectivist, we say the following. The contention that the only sensible interpretation of probability is as subjective degree of rational belief might seem to be bad news for probability statements in mathematics, which are meant to be independent of the speaker. But one way to understand the mathematician's interpretation of probability is as an objective and normative guide to what one's degree of belief should be. Suppose you need to place a competitive bid on a contract that will pay you a dollar if a certain integer n, unknown to you, is squarefree. The subjectivist agrees that your action in this situation ought to be determined by your degree of belief that n is squarefree. The mathematician argues that you should bid no higher than $6/\pi^2$ dollars, and that this is the case independent of any prior beliefs you hold (unless, of course, you have some reasonable prior beliefs about the identity of n). Furthermore, given that subjectivists usually want probabilities

[6] Easwaran [2008a] discusses the role of subjective degrees of belief in pure mathematics in connection with 'probabilistic proofs'.

to be *normative* (probabilities are said to describe how the degrees of belief of an agent rationally *ought* to be related), we suggest that this normativity often has a source outside the mind. Chance processes are one such source, which empirical sciences describe; frequencies are another, as described in number theory.

Both of the skeptics described in this section stamp their feet and insist that there is just one thing that probability can mean (though they happen to disagree about what that one thing is). Foot stamping of this sort is out of place in the problem at hand, since the fact that there are various candidate interpretations for probability has long been part of the very fiber of this philosophical problem. The frequency interpretation may be wrong as a fully general account of what probability statements always mean, or it may be right for some bodies of discourse and wrong for others, but it takes a more focused argument than the ones these skeptics produce to show that this is so. We believe that such an argument is available when it comes to many empirical probability statements, but these arguments do not touch the frequency interpretation of probabilities in number-theoretic theorems.

8. Closure Under Conjunction

When probabilities are interpreted as limits of frequencies, it can turn out that $\Pr(A \,\&\, B)$ does not have a value, even though $\Pr(A)$ and $\Pr(B)$ do. In other words, the set of propositions that have probabilities is not closed under conjunction. For example, suppose X is the Benford proposition 'n has first digit 1', which does not have a probability in the 'natural density' sense. Let A be the proposition 'n is even' and B the proposition 'n is even and has first digit 1, or n is odd and has first digit other than 1'. It is not hard to verify that both A and B have natural density $1/2$, but their conjunction is the proposition 'n is even and has first digit 1', which has no natural density.

Must a legitimate interpretation of probability entail that the set of propositions that have probabilities is closed under conjunction? We think the answer is *no*, and our reason has nothing much to do with the frequency interpretation. Pretty much anyone who believes in objective probabilities will claim that *some* propositions simply do not have objective probabilities. When probabilities are discussed in connection with empirical sciences, defenders of objective interpretations often cite Newton's laws of motion or Darwin's theory of evolution as examples; these theories may *confer* probabilities on observations, but they do not, themselves, *have* objective probabilities. Any interpretation in which some propositions fail to have probabilities is apt to deny that the set of propositions that have probabilities is closed under conjunction. Suppose that proposition X fails to have an objective probability. Now consider a fair coin which is marked 'true' on one side and 'false' on the other, and let proposition Y be 'the coin lands "true" the next time you toss it'. So $\Pr(Y) = 1/2$. Finally, let Z be the proposition 'the word on the coin that lands face up agrees with the truth value of X'. It is hard to deny that $\Pr(Z)$ should be taken to be $1/2$. For example, imagine that the truth value of proposition X was determined long ago and written on a hidden sheet of paper; then surely the probability that the coin agrees with the paper is $1/2$, independently of what is marked on the paper or how the mark was chosen. Now $\Pr(Z)$ and $\Pr(Y)$ are both $1/2$, but Y & Z is logically equivalent with 'the coin says true and X'. If two statements have probabilities only if their conjunction does too, then 'the coin says true and X' has a probability. By an exactly analogous argument, so does the statement 'the coin says false and X'. The last two quoted sentences are incompatible; so the probability of their disjunction should be the sum of the probabilities of the disjuncts (assuming finite additivity). But the disjunction is logically equivalent with X, which, by assumption, does not have a probability at all. QED.

9. Interpretations of Probability — Austere and Metaphyiscal

When empirical probability statements about coin tosses, genotypes, and half-lives are considered, actual frequency seems to be the 'least metaphysical' interpretation of probability, the interpretation most completely and directly grounded in what we observe. Actual frequencies are less metaphysical than hypothetical frequencies or propensities. Yet number theory apparently can get by with this austere concept of probability. How is that possible? In more mundane physical contexts, we are driven to use a more metaphysical concept of probability. For example, as soon as we want to say that a coin can be fair even though it is tossed an odd number of times, we are forced to abandon the actual frequency interpretation. Perhaps the reason for this is that the domains of objects explored by mathematics are already 'metaphysical', and so there is no need to supplement the metaphysics already present with a heavily metaphysical notion of probability. For example, the set of integers that number theorists consider covers all possible integers. It makes no sense to consider the integers discussed in mathematics as a sample drawn from some larger domain of integers. This corresponds to a familiar point about possible worlds: talk of necessity and possibility becomes extensional when the domain includes possible worlds.

10. Coins *versus* Numbers

We have argued that the actual-frequency interpretation and the hypothetical-frequency interpretation are both wrong for statements like 'this coin has a probability of 1/2 of landing heads when it is tossed', but that a frequency interpretation is correct for statements like 'the probability that an integer is even is 1/2'. One important difference between these two statements is that real coins get tossed only a finite number of times while there are infinitely many integers. However, frequency interpretations are wrong for coins for reasons that go beyond this simple fact. Suppose, *per impossibile*, that a fair coin is actually tossed an infinite number of times. As noted earlier, it is logically possible that the coin will land heads every time it is tossed (though this outcome has a probability of zero). But let us focus on a particular infinite sequence of heads and tails, whatever it happens to be. An infinite sequence of heads and tails can be understood in the same way as an infinite sequence of odd and even integers. Just as we can compute the limit of the frequency of even numbers in the latter, we can compute the limit of the frequency of heads in the former (if such a limit exists). But the probability thus defined with reference to this infinite sequence of heads and tails is not the same thing as the probability that the coin has of landing heads. There is no such thing as *the* sequence of heads and tails that a fair coin would have to produce. In contrast, the greater-than relation induces a unique sequence of integers. This is why coins and numbers need to be treated differently, a point that goes beyond the finite/infinite distinction.

11. Conclusion

Although the actual-frequency interpretation of probability is grossly inadequate when it comes to understanding empirical discourse about gambling devices, Mendelian reproduction, and atomic half-lives, it works just fine for the probability statements that number theorists prove like (5) and (6). A probability statement about an infinite domain is understood by taking the limit of the actual frequencies that obtain in actual finite domains of increasing size. Some of the values for these limits are settled by the natural density; these then are supplemented by Dirichlet densities. It is an open question whether further supplementations are feasible. Theorems about such probabilities are mathematical truths, and thus the idea that some probability statements are logical truths gains credence. Frequency interpretations and logical interpretations of probability are not at odds in this context. Whether all theorems about probability in number theory and in other branches of pure mathematics can be understood in this format merits further investigation.

References

Carnap, R. [1950a]: 'Empiricism, semantics, and ontology', *Revue Internationale de Philosophie* **4**, 20–40. Reprinted in *Meaning and Necessity*, pp.205–221. Chicago: University of Chicago Press, 1956.

——— [1950b]: *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Easwaran, K. [2008a]: 'Probabilistic proofs and transferability', *Philosophia Mathematica* (3) **17**, 341–362.

——— [2008b]: 'What conditional probabilities must (almost) be', unpublished.

Fitelson, B. [2007]: 'Logical foundations of evidential support', *Philosophy of Science* **73**, 500–512.

Hájek, A. [2003]: 'What conditional probabilities could not be', *Synthese* **137**, 273–323.

——— [2009]: 'Interpretations of probability', *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/probability-interpret/

Hardy, G.H., and E.M. Wright [2008]: *An Introduction to the Theory of Numbers*. 6th ed. Oxford: Oxford University Press.

Hill, T.P. [1998]: 'The first digital phenomenon', *American Scientist* **86**, 358–363.

Kolmogorov, A.N. [1950]: *Foundations of the Theory of Probability*. New York: Chelsea.

Lang, S. [1994]: *Algebraic Number Theory*. 2nd ed. Berlin: Springer-Verlag.

Popper, K. [1959]: *The Logic of Scientific Discovery*. New York: Harper and Row.

Salmon, W. [1984]: *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Skyrms, B. [1980]: *Causal Necessity*. New Haven: Yale University Press.

Sober, E. [2000]: 'Quine's two dogmas', *Proceedings of the Aristotelian Society* **74**, 237–280.

——— [2010]: 'Evolutionary theory and the reality of macro-probabilities', in E. Eells and J. Fetzer, eds, *The Place of Probability in Science*, pp. 133–162. Boston Studies in the Philosophy of Science. New York: Springer.

——— [to appear]: '*A priori* causal models of natural selection', *Australasian Journal of Philosophy*.

Venn, J. [1876]: *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability*. London: Macmillan.