

Semantics for Interval Probabilities

Henry E. Kyburg, Jr.*

The Institute for Human and Machine Intelligence
and
The University of Rochester
hkyburg@ai.uwf.edu

Abstract

We look at the several sorts of semantics that have been provided for “probability”, and explore the possibilities of generalizing them to imprecise probabilities. The *Support Set* of a statement is a useful auxiliary construct, however probability is interpreted.

Introduction

Meaning is a many faceted thing. We may want to understand the *meaning* of a term like “probability” because we need to know it to evaluate the truth of statements in which it occurs. We may want to know the meaning of the term in order to get a handle on computing the values of the function it represents. We may want to know the meaning in order to relate it to other parts of a technical vocabulary, or to explore its relations to other terms such as “entailment” or “knowledge.” In a formal context we may provide an interpretation or meaning for an expression to establish the consistency of a set of axioms containing it, or to establish the independence of some axioms from others, in short to provide a *model* of probability for one purpose or another. Or we may want to understand how to use “probability” — what is its practical import to us? What is its connection to the real world?

What follows is designed to address a number of these questions, but is primarily focussed on this last question. This is a question to which a relatively formal answer, rather than informal discussion seems appropriate. It is all very well to say that probability is a guide in life, or that we should make choices in conformity with mathematical expectation, but without some tie to the *world* it is unclear what the force of such exhortations is.

The basic idea is that of a support set: A *support set* of a probability statement is a set of models of our language that renders that probability statement true. This is so whether the probability statement is real valued, interval valued, or set valued, and whether the basic idea of the interpretation is frequency, logical

measure, or degree of belief. Of course we will have to construe “model” somewhat differently in these different cases.

Frequency Semantics for Probability

The idea of providing a formal semantics for probability is not new. Richard von Mises provided a formal model or interpretation for probability. It has been said (I have said it myself!) that his interpretation is that typically adopted by classical statisticians, but this isn’t quite true. What von Mises did was to assuage the consciences of those who wanted to associate “long run frequency” with probability. In our terms, he provided the basic idea of a formal model of an object — limiting frequency in a collective — of which the axioms of probability are true. A support model in this case is an infinite sequence satisfying von Mises’ constraints. There were difficulties concerning the irregularity requirement, which led to important ideas in computational complexity (Church 1940; Martin-Lof 1966) but basically the objects required were shown to exist as mathematical objects by Copeland (Copeland 1937).

This is a nice bit of mathematics. But nobody supposed there really were collectives in nature: infinite sequences with true randomness. Nobody ever supposed that the probability of an American male in good health of age 70 surviving to age 71 was the limiting value of a relative frequency in an actual infinite collective. The model of von Mises served to confirm the consistency of the probability axioms under the limiting frequency interpretation.

As Bertrand Russell, among others, observed, so far as empirical probability goes, this idealization is overkill. Relative frequency in a large finite actual population will do just fine as a model of the probability axioms. Furthermore, there is no doubt that these relative frequencies exist. As an interpretation of “probability” in some of its occurrences — for example those corresponding to what Pollock calls “indefinite probabilities” (Pollock 1990), or the uses that typically employ the indefinite article (the probability that *a* 70 year old man will survive...) — this relative frequency interpretation seems fine. A support model is just a finite reference class.

*This material is based upon work supported by the National Science Foundation under Grants STS 9906128 and IIS 0082928.

But this will not work for “definite” (Pollock 1990) probability statements. The only relative frequency that can apply directly to heads on the *next* toss of this particular coin, or to the survival of a *particular* 70-year-old man for a certain period, is 1 (the coin lands heads; the man survives) or 0.

Probabilities for Sentences

Carnap, among others, (Carnap 1950; Hintikka 1966) offered a very different semantics for probability. Carnap wanted to accommodate sentences like “The probability that the next toss of this coin will yield heads is 0.5,” or “The probability that Mr. Smith will survive for a year is 0.912.”

First of all, these probabilities should be construed as conditioned on “what we know;” this was recognized by Keynes (Keynes 1952). That’s easy enough to accomplish: they may be represented as the ratio of two measures on propositions: the probability of h given e , $P(h, e)$, will be just $\frac{m(h \wedge e)}{m(e)}$, where e is everything we know and $m(e) > 0$. It is thus easy enough to reduce the problem of computing probabilities to that of finding a logical measure m on propositions. In general and modern terms the measure m was thought of as defined on the models of the language. It is more convenient to take the basic measure function m as part of the model.

A support set for the probability statement “ $P(h) = c$ ” is thus a set of pairs $\langle M, m \rangle$ such that M is a model of (part of) our language, m a measure function, and $m(h) = c$; a support set of “ $P(h|e) = d$ ” is the set of pairs $\langle M, m \rangle$ such that $m(h \wedge e)/m(e) = d$. In either case there are a great many such pairs.

Carnap himself made several attempts to find a logical measure function m that would be compelling, that could be thought of as providing a *logical* semantics for probability, but he never achieved anything he regarded as satisfactory. In our terms, he was seeking a singleton support set. His initial idea was to suppose that each model of the language should carry the same weight, as suggested by an intuitive principle of indifference. This didn’t work. Assigning equal measures first to statistical structures, and then to the models that make up those structures, was more intuitive. But both Carnap and others who followed this approach have ended up characterizing the function m by means of parameters that have only intuitive justification.

An alternative is to take the measure function m to be, not logical, but psychological — to reflect the actual degrees of belief of an idealized agent. (If this is not an oxymoron!) This agent must be thought of as idealized, of course, since it is unlikely that any actual person is perfectly consistent in probabilistic terms, but that degree of approximation may be something we can live with. If we do this, we have a completely different sort of semantics for probability: one in which the *meaning* of a probability statement is to be cashed out in psychological terms — in terms of degrees of belief, rather than in terms of partial entailment.

The support sets will be similar in structure and multiplicity to those for logical probability. In particular each ideal agent may have his own belief measure.

Interval Probability

Each of these interpretations of probability can easily be extended to intervals, or, more generally, *sets* of probabilities.

Frequencies

Just as we may say that the relative frequency of heads among tosses is 0.48734, so we may say that the relative frequency of heads is between 0.450 and 0.550. Any collection of frequency functions may be taken to characterize relative frequencies in a set. Intervals are the sorts of things we can plausibly take ourselves to *know*, but there are cases in which convexity does not hold: the coin which is either biased 2:1 in favor of heads or 2:1 in favor of tails is certainly not *fair*.

If we interpret probabilities as class ratios, a model of the first statement is one in which the ratio of the cardinality of tosses that result in heads to the cardinality of tosses in general is exactly 0.48734. Even if we fix the domain of our models, there are many such models, since there are many possible arrangements of heads and tails among the individuals that make up this set. This set of models is an example of a support set of models for the statement that the relative frequency of heads among tosses is 0.48734. Whatever interpretation is given to the language, this relative frequency will hold.

The support set of the second statement is the larger set of models in which the relative frequency of heads among tosses lies between 0.450 and 0.550.

Logical and Psychological Measures

In principle the generalization from one measure to a set of measures seems elementary. C. A. B. Smith (Smith 1961; 1965), for example, took the natural path of distinguishing between the odds and agent would offer *on* an event E and the odds he would offer *against* E . This approach leads to upper and lower bounding measures on each proposition. It can be argued that these upper and lower odds must correspond to the envelop of a set of probability functions. Rational belief, as evidenced by behavior, is not convex on this model: If I will bet at no more than 1:2 on heads, and at no more than 1:2 on tails, my probability of heads is [0.33,0.66]. But I will not bet at even money.

In this case a support set for the probability of E is the set of intended models (of our language, with given empirical domain) compatible with our knowledge base, together with a measure m on that set. The (interval) value of the probability is derived from the models and a set of measure functions \mathcal{M} defined on the language: the probability of E is the set of conditional probabilities of E *given* K , where K is our background knowledge: $\{m(S \wedge K)/m(K) : m \in \mathcal{M}\}$. Behavior is or

ought to be determined by the envelop of the set of measures.

Evidential Probability

Evidential probability requires that probabilities be based on known statistical facts; it is the imprecision of our knowledge of statistical facts that renders evidential probability imprecise. Before turning to the semantics, I shall review the syntax.

Syntax

Probabilities are defined relative to a body of knowledge — a set of sentences — K . First of all we consider equivalence classes of statements of the language, where the equivalence relation is factual equivalence given K . If $K \vdash \ulcorner T \equiv S \urcorner$, then S and T belong to the same equivalence class.¹ (Note that we are not demanding that S and T be logically equivalent.)

Second, in each equivalence class we look at the statements of a particular form. We will write $\bar{\eta}$ for a sequence of variables ranging over empirical objects, $\bar{\alpha}$ and $\bar{\beta}$ for sequences of terms of any sort; all the terms in $\ulcorner \bar{\alpha} \urcorner$ are also in $\ulcorner \bar{\beta} \urcorner$. We use the Greek letters τ and ρ for formulas; τ is to suggest “target” and ρ is to suggest “reference.” It is understood that the variables free in τ and ρ are variables from the sequence $\bar{\eta}$. Furthermore $\ulcorner \rho(\bar{\beta}) \urcorner$ is ρ with its variables replaced by terms from the sequence $\bar{\beta}$, and similarly for $\ulcorner \tau(\bar{\alpha}) \urcorner$.

A potential probability for a sentence S given K arises when there is a target formula τ , a reference formula ρ , and sequences of terms $\ulcorner \bar{\alpha} \urcorner$ and $\ulcorner \bar{\beta} \urcorner$, such that:

1. $\ulcorner \tau(\bar{\alpha}) \equiv S \urcorner \in K$.
2. $\ulcorner \rho(\bar{\beta}) \urcorner \in K$.
3. For some constants p and q , $\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner \in K$.

The formula $\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner$ represents statistical knowledge in K and is to be understood as meaning that between p and q of the sequences of objects satisfying the formula ρ also satisfy τ .

For example, we could take the formula τ to be “... lands heads,” ρ to be “... is tossed,” and p and q to be 0.48 and 0.52: “ $\%x(x \text{ lands heads}, x \text{ is tossed}, 0.48, 0.52)$.” Or to say that the relation of being longer than is almost always transitive we could write “ $\%x, y, z(L(x, z), L(x, y) \wedge L(y, z), 0.9, 1.0)$.”

We get from potential probability intervals to actual ones with the help of three syntactic principles: *richness*, *specificity* and *precision*. These are essentially principles for *disregarding* potential reference formulas ρ . There may well be a number of distinct potential probabilities left; the *cover* of these remaining intervals is what we take to be the evidential probability associated with the equivalence class of S .

¹We follow Quine (Quine 1041) in using corners to name types of expressions.

Semantics

In view of the fact that our object language contains statements mentioning real numbers, models are best construed as having two components: a domain of mathematical objects and a distinct domain of empirical objects \mathcal{D}_e . We stipulate that the domain of empirical objects be finite.

First we require semantics for statistical statements — statements of the form $\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner$. To this end we introduce the notion of a *satisfaction set* of a formula ϕ in a model m . Intuitively, this is the set of empirical objects that satisfy the formula ϕ .

$SS_m(\phi) = \{ \langle a_1, \dots, a_n \rangle \in \mathcal{D}_e^n : v_{x_1, \dots, x_n}^{a_1, \dots, a_n} \text{ satisfies } \phi \}$, where ϕ contains just the n free variables x_1, \dots, x_n , and $v_{x_1, \dots, x_n}^{a_1, \dots, a_n}$ is any variable assignment that assigns a_1 to x_1, \dots, a_n to x_n .

With the help of this notion, we can define truth in a model m for statistical statements:

$\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner$ is true in m if and only if $|SS_m(\rho)| > 0$ and $p \leq |SS_m(\tau \wedge \rho)| / |SS_m(\rho)| \leq q$.

The statistical information that underlies probability intervals can be reflected in model theoretic relations concerning the ratios of satisfaction sets and the structure of reference sets.

For the sentence S and the background K the i 'th *support set* of models, Λ_i , is the set of models in which $\ulcorner S \equiv \tau_i(\alpha_i) \urcorner$, $\ulcorner \rho_i(\alpha_i) \urcorner$, and $\ulcorner \% \bar{\eta}(\tau_i, \rho_i, p_i, q_i) \urcorner$ are all in K and ρ_i has survived the application of our three principles. $\ulcorner Prob(S, K) = [p, q] \urcorner$ is true if $[p, q]$ is the cover of the intervals mentioned in the support set: p and q bound the frequency of truth of S in the set of support models.

General Support Sets

As we noticed earlier, almost any standard interpretation of probability can be generalized to allow for sets of probabilities or for probability bounds, or both. Let us see how the semantics for evidential probability can reflect these interpretations.

Frequencies

It follows from the fact that reference classes are finite, that there is exactly one true relative frequency of τ 's among ρ 's in any model. The formula $\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner$ will be true in a model just in case this relative frequency falls between p and q in that model.

We may be interested not only in the *general* statistical statement, but also in its, shall we say, instantiations. Thus not only do we want to specify truth conditions for $\ulcorner \% \bar{\eta}(\tau, \rho, p, q) \urcorner$, but also for “the probability of $\ulcorner \tau(\bar{\alpha}) \urcorner$ is $[p, q]$ ” for certain sequences of terms $\bar{\alpha}$.

Typical cases in which we want such expressions to inherit the intervals underlying the general case arise in physics (particularly statistical mechanics, if we are content to let the number of particles be finite), in well regulated gambling apparatus (provided that $\bar{\alpha}$ refers to

a *future* event), and similar cases in which the events are, or are assumed to be, outside of human control.

But demographic cases are also often taken to fall into this class. The insurance company assumes that the probability that Mr. Smith will survive the coming year is the (approximate) frequency in the reference class into which his characteristics place him. It is not that his survival is beyond human control; in a rather grim sense it is well within human control to determine that he does not survive the year.

In each case, however, the probability is the relative frequency of truth of $\lceil \tau(\bar{\alpha}) \rceil$ in the support set of *relevant* models, where the support set of relevant models is determined by ρ (the reference class is always mentioned or implicit in cases amenable to frequency interpretation), τ , and $\bar{\alpha}$. In effect, the set of models is generated by allowing $\bar{\eta}$ to range over the entire interpretation of ρ . The interval probability of $\lceil \tau(\bar{\alpha}) \rceil$ is just the range of ratios of $SS_m \lceil \tau(\bar{\eta}) \wedge \rho(\bar{\eta}) \rceil$ to $SS_m \lceil \rho(\bar{\eta}) \rceil$ in models m .

Example: “The probability of survival for another year of a 40 year old man is [0.92,0.95].” Let \mathcal{D} be the domain of our model. In each model the satisfaction set of “ x is a 40 year old man” has a certain cardinality, as does the satisfaction set of “ x is a 40 year old man who survives for another year.” In order to make the probability statement true, the ratio of these two cardinalities must lie in [0.92,0.95]. The support set of models for this statement, construed in terms of frequencies, is just the set of models in which the corresponding statistical statement is true.

More interesting is the fact that “the probability of survival for another year of Charles, a 40 year old man, is [0.92,0.95]” has exactly the same support set on the frequency interpretation, when what we suppose ourselves to know about Charles is exactly that he is a 40 year old man.

Logical Width

Let us follow the trail blazed by Carnap (Carnap 1950) and Jaynes (Jaynes 1958) and suppose that there is a syntactically determined measure defined for each (possible) sentence in our formal language, at least within a context. This approach generalizes naturally to indeterminate probabilities: we suppose that there is a set of distributions or logical measures, rather than one. How can we spell out the truth conditions of “the probability of S is the interval $[p, q]$?”

In the simple case, the probability of S will be the sum of the measures on the intended models in which S is true. When there is a set of measures on these models, this no longer holds. For example, if we are modelling the rolls of a die, and the measures we assign to the six models of the outcome are 0.1,0.1,0.1,0.1, p ,0.6- p , for $0.0 \leq p \leq 0.6$, and S is “lands five or six” the looseness of the logical measure for “lands five” is irrelevant.

The support set of models for S consists of intended

models paired with measures. In this case it is the set of intended models of S , paired with the singleton set of measures that assigns measure 0.6 to S .

Coconditional Probability is more interesting. For any S there will always be a pair of statements of the form $\tau(\bar{\alpha})$ and $\rho(\bar{\beta})$ where $\tau(\bar{\alpha})$ is known to be equivalent to S and $\rho(\bar{\beta})$ is part of our data. (We assume the full power of first order logic.) Any such $\rho(\bar{\beta})$ may be taken to determine a potential support set Λ consisting of the set of models in which $\rho(\bar{\beta})$ is true, paired with a *set* of measures \mathcal{M} determining the truth frequency of $\lceil \tau(\bar{\alpha}) \rceil$ in Λ . (In evidential probability the set of models is constrained also by the truth of some statistical statement.)

We may thus make a connection between evidential and logical-measure probability. The support set of a logical measure probability will always be a subset of *each* of the support sets of evidential probability. The connection does us little good, however, since the measure on the truth of S in a set of models need not at all constrain the measure on the truth of S in a subset of those models.

Example:

The probability that Lightning will win the 2:00 race, given our total knowledge base K is the set of conditional probabilities obtained by conditioning that event on K , using each of the set of initial measures with which we start.

But K includes many facts that I regard as irrelevant. Let K^* be a minimal set of evidence statements such that we obtain the same conditional probabilities for Lightning’s success, and such that the deletion of any subset of K^* would alter the conditional probability. The models and measures of each such set of statements is also a support set for that probability. Note that each supports the same probability range. But also note that there is no automatic way in which to take account of a *new* item of evidence.

Subjective Belief

The value of a logical probability, whether definite or indefinite, is intended to be legislative or normative for rational belief; the *truth* or *falsity* of probability assertions is taken to be a matter of logic, and not a matter of psychology. The difficulty raised by Frank Ramsey (Ramsey 1931) for the system of Keynes (Keynes 1952) was the difficulty of knowing when the logical system was being obeyed. But then Ramsey went on to argue that all we can ask of a person is that his degrees of belief conform to the probability calculus, not that they be otherwise constrained, where “degrees of belief” are taken to be revealed behaviorally.

If probabilities are idealized degrees of belief, then the semantics must refer to an idealized psychological domain. Part of the idealization is that the objects of belief are taken to be represented by the sentences of a formal language. We may also abandon the idea that

beliefs are real-valued. We retain the idea of logical omniscience.

An idealized state of mind can then be represented by the intended models of the language, together with a degree-of-belief function B taking real values. Since we have finitized and sanitized things, we are in the same position we were in the case of logical measures. Despite the fact that we interpret probabilities as psychological rather than logical, the support sets of $P(S) = [p, q]$ are still ordered pairs, consisting of intended models $M(S)$ paired with a set of “degree” of belief functions \mathcal{B} defined over that set. In this case there may be good arguments for regarding \mathcal{B} as a convex set.

Example:

I think the probability that Samuels will be elected mayor is the interval $[0.30, 0.40]$. This interval is the result of conditioning on everything in my background knowledge K . But it may well be that I don't regard anything but E_1 and E_2 as relevant. Thus the support set for this probability may be simplified to the much larger set of models making E_1 and E_2 true, together with the set of measures $\{B(S \wedge E_2 \wedge E_1)/B(E_2 \wedge E_1) : B \in \mathcal{B}\}$.

Again, there is no built in procedure for taking account of a new item of relevant evidence E_3 .

Discussion

Since each support set is a superset of K , our knowledge base, we know that the intersection of these support sets is nonempty. However, just as we know that in the “real” world, S is true, or in the “real” world S is false, so too in the collection of models that render our total knowledge base true, the relative frequency of models in which S is true has some determinate value (of course); but we don't know what it is.

There are two responses. One is to make some “assumptions”—to suppose that we know things that in fact we don't know. This is what happens when we take subjective probability as an “estimate” of relative frequency, or adopt a logical measure function. The other response is to disregard some of the things we do know. This is what happens when we regard the next toss of a coin as “merely” a toss, and not a toss performed by Jane on Tuesday.

Evidential probabilities are based on frequencies that are known to hold in the real world — it is *objective*. The semantics we have just outlined illustrate how that comes to pass. What allows this to happen is that in the interest of objectivity we ignore some of the data we have. We are not “assuming” extra knowledge, but rather *disregarding* irrelevant knowledge; thus the set of models in which the relevant frequency is calculated is a *superset* of the models of our background knowledge.

Note that in many cases there is essentially only one support set Λ_i in the sense that if Λ_j is another support set, then the intervals mentioned in the component statistical statements are the same. This is true, for example, for the outcomes of well constructing gambling

apparatus, for measurement error, etc. In these well regulated and well understood cases, there is essentially only one reference class, and the known limits on the relative frequency in that reference class determine the probability.

Conclusion

The semantics we have generated for evidential probability, in which probabilities are intervals, admits of variations corresponding to frequency interpretations of probability, logical measure interpretations, and personalistic degree of belief interpretations. The general form of the semantics, expressed in terms of support sets, seems quite useful.

References

- Rudolf Carnap. *The Logical Foundations of Probability*. University of Chicago Press, Chicago, 1950.
- Alonzo Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46:130–135, 1940.
- Arthur H. Sr. Copeland. Consistency of the conditions determining kollektivs. *Transactions of the American Mathematical Society*, 42:333–357, 1937.
- Jaakko Hintikka. A two-dimensional continuum of inductive logic. In Jaakko Hintikka and Patrick Suppes, editors, *Aspects of Inductive Logic*, pages 113–132. North Holland, Amsterdam, 1966.
- E. T. Jaynes. Probability theory in science and engineering. *Colloquium Lectures in Pure and Applied Science*, 4:152–187, 1958.
- John Maynard Keynes. *A Treatise on Probability*. Macmillan and Co., London, 1952.
- Per Martin-Lof. The definition of random sequences. *Information and Control*, 9:602–619, 1966.
- John L. Pollock. *Nomic Probability and the Foundations of Induction*. Oxford University Press, New York, 1990.
- W. V. O. Quine. *Mathematical Logic*. Harvard University Press, Cambridge, 1041.
- Frank P. Ramsey. Probability and partial belief. In *The Foundations of Mathematics*, volume London, pages 256–257. The Humanities Press, 1931.
- C.A.B. Smith. Consistency in statistical inference and decision. *the Journal of the Royal Statistical Society B*, 23:1–37., 1961.
- C.A.B. Smith. Personal probability and statistical analysis. *the Journal of the Royal Statistical Society A*, 128, pages 469–499, 1965.