OBJECTIVE PROBABILITIES

Henry E. Kyburg
Philosophy and Computer Science
University of Rochester

ABSTRACT

The common distinction between probabilities that can be based on frequencies known to hold in a sequence of repeatable events, and probabilities that concern unique events, and that therefore must be based on subjective opinion, is argued to be misguided. All events are in some relevant sense "unique", and, more importantly, all events can in a relevant sense be placed in classes of similar events. A formal calculus is described for accomplishing this in relatively simple but useful cases. *

A. Statistics and Unique Events

Almost everyone will agree that when our background in statistical knowledge is extensive enough, and when the case with which we are concerned is a "repeatable" event, then objective probabilities are appropriate, and these are the probabilities that should enter into the computation of expectation and into our decision theory. A great many people will also agree that there is another whole class of cases, in which we are concerned with unique events, in which we lack statistical knowledge, and for which we must turn to subjective probability or one of its surrogates. I propose here to argue against this distinction. Of course it is easy enough to argue this way in a purely philosophical vein: every event must be unique --it has its own spatio-temporal locus; and every event must belong to some class of events about which, in principle, we could have statistical knowledge. But this is not my point. My point is that from a down-to-earth practical point of view, from the point of view that seeks to compute probabilities and expectations for making decisions, the distinction between 'repeatable' and 'unique* events is not only untenable, but seriously misleading.

B. Kinds of Cases

Let us consider some examples of these alleged distinctions. Consider the toss of a coin. There is a classical 'repeatable' event: not only can we toss a coin over and over again; coins have been tossed over and over again, and in the experience of each of us there is a large data-base of results of coin tosses (or an impressionistic resume of such a data-base). And we have physical grounds (i.e., grounds stemming from the laws of physics for thinking that coins land heads about half the time. And so we can regard that toss of that coin as a member of a class of tosses, of which we have reason to believe that half yield heads. (Alternatively, we might regard that toss as a kind of trial that has a propensity of a half to yield heads.

Now of course a particular toss, at a particular time and place, cannot be repeated. We all know that. But the event can be repeated 'in all relevant respects*. We don't have to make the toss at the same time or the same place; we don't have to use the same coin; we don't have to use the same kind of coin; we don't have to flip it in any particular way.

Consider another kind of case. I want the probability that my friend Sam will be at home tonight after supper. I know that on some week nights he goes to the movies. This is unlike the coin in that I can specify some of the factors that lead him to go out or to stay home. Thus I might know that he likes westerns ~ so if there is a good western in town, he will be more likely to be at the movies. On the other hand, I might know that he is very conscientious about studying for his chemistry examinations - so if there were to be a chemistry examination tomorrow, Sam would be most likely to be at home. It is hard to specify a "repeatable" event. In short this seems like a perfect case for subjective probability. But if I know Sam well, I will have some basis for knowing how often, in general, he goes out on week-nights. My knowledge is neither so precise nor secure as my knowledge of the coin, but surely it is not non-existent.

But we must beware of allowing the variety of bur knowledge about Sam to serve as an excuse for guessing wildly. Analogously, if we were to have detailed and microscopic data concerning the coin toss, we could perhaps predict with a better than 50% success rate. This possibility should not be allowed to undermine our sensible tendency to assign a probability of a half to the occurrence of heads on the specified toss when we lack that microscopic data.

* This work has been supported in part by the Signals Warfare Center.

Finally, there are some circumstances under which my probability concerning Sam's being at home is just as exact as my probability concerning the flip of a coin. For example, I may know that he decided whether or not to go to the movies by flipping an ordinary coin. Suppose in addition that I know that he went out if and only if the coin landed heads. Then the probability that Sam went to the movies is 0.5. This brings out an important point that I shall enshrine as an axiom:

AI *If S and T are known to have the same truth-value, then they have the same probability.*

This axiom does not require that S and *T* be equivalent in any strong sense; all that is required is that we know that they have the same truth value.

This axiom already undermines the argument for subjectivity based on uniqueness. It is asked, "How can you find an objective probability for the event of a New York nuclear power plant suffering a meltdown, when there is no class of instances to generalize from: there is only one New York nuclear power plant, its design is unique, etc." The answer is that we need not be concerned about the frequency of failure in plants of such and such design, but rather can transform that sentence into one having the same truth value, to which our statistical knowledge is applicable. (The plant will fail if and only if gate valve #1 fails or gate valve #2 fails...)

Here is another example. I hold in my hand a newly minted coin. I will toss it once, and then melt it down. What is the probability that this coin will yield heads when tossed? The relative frequency among tosses of this coin is 0 or 1 - we have no statistical knowledge of the behavior of this coin. But we know that the toss of this coin I am about to perform will yield heads iff the next toss of a coin yields heads -- and for tosses of coins in general we have lots of statistical evidence.

C. <u>A Simple System</u>

Here is a very simple example of how objective probability can be applied to "unique" events. It is essentially due to Reichenbach (1949).

Let $R = \{r_1,...,r_2\}$ be a finite set of potential reference classes; let $P = \{P_1,...,P_2\}$ be a finite set of properties (including such properties as being a member of a particular class), and let $I = \{i_1,...,i_2\}$ be a set of distinct individuals. We can define a language on this basis in the usual way.

Add to this language enough mathematics to do statistics, and define an <u>*item of possible statistical knowledge*</u> to be a sentence of the syntactical form "$\%(r_j,P_s)$", *which we read: the proportion of objects in the reference class* r, *that have the property* P_s *is x.* Proportions satisfy the classical probability calculus.

Let a body of knowledge *K* be a set of sentences. We impose few restrictions on K. We want it to be consistent in the sense that there should be no sentence S in *K* for which _ S is also in *K.* We want some logical truths in *K.* We want the items of possible statistical knowledge to be consistent statements concerning the relations of the possible reference classes and properties. Finally, we want the (finite number of) sentences of the form "$P_z(i_y)$" together with "$0=0$" and "$0=1$" to generate a partition of all the sentences of the language under the relation of *being known to have the same truth value.*

These constraints are embodied in the following axioms:

A2.1 *R* is closed under intersection

A2.2 *If i and j are distinct, r_i and r_j are in R, then $r_i = r_j$ is not provable.*

A2.3 *If* $\vdash r_i \subset r_j$ *then* "$r_i \subset r_j$" $\in K$.

A3.1 *P is closed under conjunction and atomic negation.*

A3.2 *If i and j are distinct, P_i and P_j are in P, then* $\forall x(P_i(x) \Leftrightarrow P_j(x))$ *is not provable.*

A3.3 *If* $\vdash \forall x(P_i(x) \Rightarrow P_j(x))$ *then* "$\forall x(P_i(x) \Rightarrow P_j(x))$" $\in K$.

A4 *If* "$i_x \in r_y$" $\in K$ *and* "$i_x \in r_w$" $\in K$, *the* "$i_x \in r_y \cap r_w$" $\in K$.

A5.1 *If* "$S_1 \Leftrightarrow S_2$" $\in K$, *then* "$S_2 \Leftrightarrow S_1$" $\in K$.

A5.2 *If* "$S_1 \Leftrightarrow S_2$" $\in K$, *and* "$S_2 \Leftrightarrow S_3$" $\in K$ *then* "$S_1 \Leftrightarrow S_3$" $\in K$.

A5.3 "$S_1 \Leftrightarrow S_1$" $\in K$.

A6 *For every non-mathematical** sentence S in L, there exists a P_y and exactly one i_x such that* "$S \Leftrightarrow P_y(i_x)$" $\in K$.

A7 *There exists a model of the sentences in K, with "$\%(X,Y)$" construed as "the proportion of X's that are Y's".*

We can now define the probability of a sentence *S* relative to a body of knowledge *K* to be *x* just in case *S* is known in *K* to be equivalent to a sentence of the form "$P_z(i_y)$" -- this is just to say that the biconditional "$S \Leftrightarrow P_z(i_y)$" is in *K* -- and for some reference class r_w to which i_y is known to belong, "$\%(r_w,P_z) = x$" is in *K*, *and, finally if* $r_{w'}$ is another reference class to which i_y is known to belong, and "$\%(r_{w'},P_z) <> x$" is in *K*, then it is known that r_w is included in $r_{w'}$. Formally,

D1 *Prob(S,K)* $= x$ *iff there are P_z, i_y, and r_w such that*

(1) "$S \Leftrightarrow P_z(i_y)$" *is in K.*

(2) "$i_y \in r_w$" $\in K$.

(3) "$\%(r_w,P_z) = x$" $\in K$

(4) If "$i_y \in r_{w'}$" $\in K$, *and* "$\%(r_{w'},P_z) <> x$" $\in K$, *then* "$r_w \subset r_{w'}$" $\in K$.

Thus r_w is the <u>smallest</u> reference class about which we have statistical information to which iy is known to belong. This is essentially Reichenbacns idea, except for the addition of axiom A I .

We can generate the probability more clearly by Eutting the fourth condition as a constraint on a table, Eet the first column of the table contain a list of all the reference classes r_w to which i_y is known to belong. Let the second column contain the value of $x(r_w)$ from the corresponding item of statistical knowledge:

**A non-mathematical statement is one whose truth value depends <u>only</u> on empirical facts.

"$\%(r_w,P_z) = x(r_w)$." **Work down the table, deleting every row that fails condition (4) (Rule: if $x(r_w) <> x(r_{w'})$, delete both rows unless "$r_w \subset r_{w'}$" is in K. There may be several rows left, but they will all mention the same value of x. There may be no rows left.**

D. Limitations

This approach deals perfectly reasonably with tosses of coins and the like. It also does what we want for Sam, in the case in which he decides whether or not to go to the movies by tossing a coin. But it has serious drawbacks. It fails to provide for the case in which we get the probability of Sam going to the movies from a limited statistical basis. It gives us no probability at all when we know of iy that it belongs to two reference classes, and our knowledge about those reference classes doesn't agree, and we don't know that either reference class is included in the other.

The remedy is simple and obvious, but it entails considerable complication. We allow items of possible statistical knowledge to embody approximate knowledge. Let us write:

"$\%(r_w,P_z) \in [x_1,x_2]$"

to mean that the proportion of objects in the reference class r_w having the property P_z is in the closed interval $[x_1,x_2]$.

Suddenly we have statistical knokwledge about every property and every reference class: at the very least we will know that the proportion lies in *[O,lJ.* And now what do we mean by <>? These changes work: Say that two intervals "differ" if neither is included in the other, and rewrite (4) to say that if r_w and r_w'differ, then r_w is known to be included in r_w\

(4') If "$i_y \in r_w$" is in K and "$\%(r_{w'},P_z) \in [x_1',x_2']$" is in K, and $[x_1',x_2']$ differs from $[x_1,x_2]$ then "$r_w \subset r_{w'}$" is in K.

And. finally, we must include another clause to single out for us the most informative interval that is not ruled out by conflict with another interval:

(5) ifr_w^f has not been eliminated as a possible reference class by the earlier conditions, then the interval corresponding to r_w is a subinterval of the interval corresponding to $r_{w'}$.

This new definition of probability is still limited - it turns out that we would like two other relations, in addition to the subset relation, to excuse "difference". (One is a gubsample relation dual to the subreference class relation. The other is a cross-product relation that accounts for Conditionalization in the presence of background knowledge.) And we would like to be able to consider equivalence to statements concerning several different individuals (Kyburg, 1985). But it is already quite powerful, and it has some rattier interesting properties:

(1) All probabilities are objective, in the sense that each proEability is based on empirical knowledge about frequencies or chances in the world.

(2) Every statement in the language has a probability; there is no distinction Detween statements concerning "re pea table" events and statements concerning "unique" events.

(3) No a priori probabilities are required; all probabilities can be based on experience. (But how they can be so based is another story.)

E. Conclusion

Probabilistic knowledge may be regarded as all of a piece. There is no need to distinguish between statistical" probabilities that have objective warrant in the world and "subjective" probabilities that merely reflect our subjective feelings. When we apply our knowledge of statistical facts to individual cases, it is the probability of a unique event that is at issue. When we offer a "subjective" probability for a unique event, it is, if it has any epistemological justification at all, based on some (possible approximate) statistical knowledge. The difference between the two cases lies in the fact that in the former case it is easy to specify the reference class -- it may even be built into tne problem through the use of the indefinite articles "a" and "an" -- and in the latter case, it may be quite difficult to put your finger on the reference class. But this is a difference of degree, and not of kind. The procedures suggested here (and in Kyburg 1985) can render both kinds computable within quite rich languages.

References;

Kyburg, H. "The Reference Class," Philosophy of Science 50 (1983) 374-397.

Reichenbach, H. Theory of Probability, University of California Press, Berkely and Los Angeles, 1949