Evidential Probability

Henry E. Kyburg, Jr.
Computer Science and Philosophy
University of Rochester
kyburg@rochester.cs.edu

Abstract*

We characterize and illustrate *evidential probability* — an interval-valued measure of uncertainty that applies to sentences. Evidential probability is also taken to be relative to a body of knowledge or a database representing the data on which judgments of uncertainty are to be made. Every statement is equivalent to some statement to which statistics in the database are relevant. The problem is to find the correct statement; this is the problem of finding the correct reference class for a statement. A set of rules for determining the correct reference class is offered.

1 Kinds of Probability

Most scientists feel most comfortable with some kind of frequency or empirical interpretation of probability [Wilson, 19521. It is not merely that this way of looking at probability seems to have a comfortable history in casinos, but that it is what physicists employ in their theoretical work in quantum mechanics as well as in statistical mechanics. It is also what is natural to have in mind in dealing with "random error" in measurement. Indeed, it is the basic idea that is called on when we need to go from the "random errors" in a number of individual measurements of different quantities, to the "inferred error" in a quantity that is taken as a function of those measured quantities. This is a relatively uncontroversial framework for "combining" probabilities.

In designing expert systems, however, what we have to work with are often the mere opinions of experts. It is said that "we have no statistics." (McCarthy and Hayes, 1969) Every event that we need to concern ourselves with, after all, is "unique." (That's the nature of events!) And there is a well developed interpretation of probability as subjective: the view developed by L. J. Savage 11954], F. P. Ramsey [1931], and Bruno de Finetti [1937]. This approach to probability has been widely adopted in business schools, and is accepted by a significant minority of statisticians [Raiffa and Schlaifer, 1961].

Prima facie, the problem with subjective probability is that t is subjective; the founding fathers just cited make crystal clear that that is exactly what they intended: one man's coherent opinion is, formally, as good as another's. Ramsey was self-consciously clear about this: "Ia man's] original expectations may within the limits of consistency be any he likes..." (p. 189)

In contrast to both of these views, *evidential probability* both applies to unique events, and is solidly based on statistical evidence. This is not a new interpretation; it is a philosophical interpretation that has been around for a number of years as "epistemological probability." The basic idea appeared in [Kyburg, 1961]. The following sections will (a) detail the difficulties of applying probabilities, construed as empirical frequencies, to specific situations, (b) illustrate further difficulties with (strictly) subjective probability, (c) illustrate the application of evidential probability in simple contexts, and (d) argue that the scope of evidential probability is in fact wider than one might at first suppose. We will, finally, (e), describe the relation between evidential probability and belief functions, subjective probability, and frequency probability.

2 Frequencies Are No Help.

Consider a circumstance in which we ordinarily take ourselves to have unbounded and dependable statistical information: a well run casino. We suppose ourselves to know, with very little error, the relative frequency with which a die yields six: 1/6. In placing a bet, however, we are concerned with a much more specific situation than "a roll of the die." It is a roll at casino X. But we don't have any statistics about casino X — only vaguely about dice in general. It is a roll by person Y; but we certainly don't have any statistical information about his frequency of sevens. And it is a roll made at 12:34 AM, on date D. If we knew the relative frequency of sevens in that set of rolls of cardinality one, we'd be in good shape; but then we would have no call to use probability.

Of course, what we do is to suppose that the general frequency can be applied to the specific instance. In this kind of case that makes good sense. There are only two circumstances in which one can (ordinarily) be with regard to the roll made at 12:34 AM on date D. Either one knows how it came out (six or not) or one knows nothing that would interfere

Research Contributing to this paper has been supported by the Signals Warfare Center of the U. S. Army.

with the use of the long run general frequency.

There are other situations one can consider. Thus, the following day, I and my fellow croupiers, having access both to the summary statistics of the previous night's outcomes, and the detailed histories, may bet on various rolls. Thus we may bet on whether or not the roll made at 12:34 AM on date D came out six. Knowing that only 1/7 of the rolls at that table on date D came out six, we will bet at odds of 1:6 rather than odds of 1:5. But this is clearly an abnormal situation.

Perhaps, as scientists, we would prefer to deal with something more important and less hypothetical, such as the distribution of errors of measurement. But even in such familiar contexts identifying probability and frequency is not quite right. *(Which* frequency?) It is one thing, in a general context, to identify the probability that a measurement will be in error by three standard deviations with the long-run frequency of an error of that magnitude. It is quite another to attempt to identify the probability that a *specific* measurement was in error by that amount. That measurement either was or was not in error by three standard deviations: the frequency is 0 or 1.

If we impute to the instance the general frequency, we had better take into account what we know of the magnitude measured, the results of other measurements of that magnitude, the results of other applications of the measuring device, the results of calibration, perhaps even the "personal equation" (discovered in 18'th century astronomy). If a measurement yields the number 4.60, then, under some circumstances the probability that the magnitude being measured lies between 4.5 and 4.7 is 0.95, But suppose that measurement is known to be one of eight measurements, the average of which is 4.50. Then clearly we would NOT want to say that the probability that the magnitude being measured lies between 4.5 and 9.0 is 95. Something more than known frequencies must be invoked. Inference from the performance of measurements is non-monotonic.

The defect of frequencies, considered by themselves, for dealing with uncertainty is even more clearly brought out by the circumstances of a hypothetical insurance company. Insurance companies collect enormous amounts of data. Suppose we want to set a fair premium for a term life insurance policy for Mr. Y. Mr. Y is a white American male of age 42, and we have statistics on the mortality of white American males of age 42.

But we give Mr. Y a questionnaire, and discover that both of his parents, and one of his sisters, died of heart disease at a very early age. Clearly this puts him in a special group: those who have a family history of heart disease; and we have statistics for mortality among that group.

On the other hand, he is a professor, and our data regarding professors suggests that they tend to suffer less from heart disease than most people.

And so on. The problem is not that, once having decided to base our uncertainties on statistics, there aren't enough statistics, but that there are too many. Mr. Y belongs to many classes concerning which we can reasonably assume ourselves to have statistical knowledge, and they all have differing and conflicting inferred relative frequencies of death.

Ultimately, of course, Mr. Y is unique: the intersection of the classes to which he belongs, about which we may even have historical data, and be in a position to make statistical inferences, is occupied by only one person, Mr. Y, and the incidence of mortality over the next year is either 0 or 1.

The philosopher Hans Reichenbach explored this problem long ago [Reichenbach, 19491. His recommendation was to adopt as a reference class determining one's assessment of likelihood the "narrowest" reference class about which one has "adequate" statistics. But this recommendation suffers from a number of defects. We may have statistical information about professors and about people with a certain kind of family history, without having any statistical information about professors who have that kind of family history. And there are clearly other considerations than "narrowness" that can be raised.

3 A Judgment Call?

In the final analysis, it is perhaps just a judgement call, we'll be told. (See, for example, Cheeseman, [1989].) Why should this be unsatisfactory? Because it leaves us with no way of adjudicating conflicting judgments. Maybe there is no way of settling such differences, perhaps it is as Ramsey suggested long ago. But if an agent's original expectations may be any he likes, so, even if he updates his beliefs by conditionalization, may his final expectations be anything at all.

Of course, people do tend to agree. They tend to agree about the likelihood of a seven on the roll of a die made at 12:34 AM on date D. Is this agreement just a reflection of good will and social skills? Or is it a result of intimidation? Or of a failure to reflect? The trouble with conventions is not that they are all wrong, but there is no basis for evaluating them, for weighing one against another.

Ah, but how about success rate? The winners must have been right! Not so; perhaps the winners were merely lucky. We do not automatically say of a man who bets on the long shot for no reason, and wins, that he was "right." The problem of evaluating uncertainty reemerges: we need to be able to assess the likelihood of the proposition that his success was reasonably to have been expected, so that we know whether to emulate him in the future.

And then there are the related problems of consistency and updating. According to the standard subjectivist view, we are to consult our opinions (judgments) and then, as evidence comes in, modify them by conditionalization. But our initial judgments are not likely to satisfy the constraints of the probability calculus: we will have to adjust them to make them come out right. This adjustment does not take place by conditionalization. We find some judgments "about which we feel sure/' to use a phrase of Savage's (1954), and adjust the others to match.

But suppose we begin with a set of judgments we think coherent, get some evidence, conditionalize, and arrive at a set of judgments we no longer find plausible. No problem. We need merely go back and (as before, on the same basis) alter our original judgments until everything comes out right.

Is this a parody? Perhaps. But it is also, from the subjective point of view, a rationally permissible scenario. We look to probability to guide our degrees of belief, our assessments of uncertainty; and what happens is that we end up using our intuitive assessments of likelihood and plausibility to control our imputations of probability.

4 Evidential Probability.

It is precisely this tension between what we know of statistics — a fair amount, by any non-skeptical standard — and the need to adjudicate among known statistics in order to achieve an assessment of the particular case that confronts us that constitutes our problem.

As we have seen, the problem with making frequencies do the work is that there are generally too many. The problem is not a lack of frequency information, but rather a surfeit of information (not merely a surfeit of observed data, but a surfeit of the information we can infer from it!). This becomes even more apparent when we note that every probabilistic measure of uncertainty has the property that if two statements, *S* and T, are known even to have the same truth value (and a fortiori if they are logically equivalent), they must be assigned the same probability.

Demonstration: Show $P(S \mid K\&(S \equiv T)) = P(T \mid K\&(S \equiv T))$. Consider $P(S\&T \mid K\&(S \equiv T)) = P(S \mid K\&(S \equiv T))^{*}P(T \mid K\&(S \equiv T)\&S) = P(T \mid K\& (S \equiv T))^{*}P(S \mid K\&(S \equiv T)\&T)$ But $P(S \mid K\&(S \equiv T)\&S) = P(T \mid K\&(S \equiv T)\&S) = 1$, so $P(S \mid K\& (S \equiv T)) = P(T \mid K\&(S \equiv T))$.

Thus the probability of heads on the next toss of this coin, if it is to be based on everything we know about this coin, must reflect the knowledge we have about coins in general, about physical systems of a certain character, about 1968 coins, about 1968 quarters, about 1968 quarters being tossed in 1990, etc. That is: we know that the next toss of a (spatio-temporally identified) 1968 quarter lands heads if and only if the next toss of a (spatio-temporally identified) coin lands heads if and only if the next toss of this coin I hold now in my hand lands heads. I have little *direct* statistical knowledge about this coin, but I have a lot of knowledge about coins in general (both folk knowledge and the results of experiments) and (on a theoretical basis) about physical systems in which a small perturbation can lead to a qualitative difference.

It follows from this principle of equivalence (sentences known to have the same truth value should have the same probability) combined with the principle of total evidence (we should take account of all the evidence we have), that if we know (of a certain object) that it belongs to a collection C of classes, the probability that it has a certain property should reflect what we know about the incidence of that property in those classes.

5 Extended Example.

Here is a table that might have resulted from a survey of beef cattle reporting whether they develop founder given *(a)* worming medicine, (b) implant, (c) louse powder, *(d)* soybean supplement, or *(e)* corn supplement, or any combination of these. The table does *not* give information regarding the effect of the lack of these features: thus *abcd* represents both those falling in class *abcd* who are given corn supplement, and those falling in that class who are not given corn supplement.

The last two columns give .90 confidence level intervals for founder, based on the evidence given in the table, using standard forms of statistical inference.

	n	founder	fl	fu
u	10000	2622	0.2493	0.2751
a	7600	1917	0.2374	0.267
b	6400	1632	0.2389	0.2711
c	4800	1225	0.2366	0.2738
d	3750	967	0.2368	0.2789
e	5400	1292	0.2217	0.2568
ab	5250	1267	0.2235	0.2591
ac	3500	855	0.2225	0.2661
ad	2500	602	0.215	0.2666
ae	4150	942	0.207	0.247
bc	2250	560	0.2217	0.2761
bd	2300	577	0.224	0.2778
be	3050	702	0.2068	0.2535
cd	900	250	0.2348	0.3208
ce	3200	710	0.1991	0.2447
de	1200	322	0.2311	0.3056
abc	1650	390	0.2046	0.2681

abd	2100	502	02109	0.2672
abe	2900	642	0.1974	0.2453
acd	400	100	0.1855	0.3145
ace	2500	490	0.1702	0.2218
ade	200	52	0.1688	0.3512
bcd	0	0	0	0
bce	1600	340	0.1803	0.2448
bde	0	0	0	0
cde	500	150	02423	0.3577
abcd	0	0	0	0
abce	1500	290	0.16	0.2266
bcde	0	0	0	0
aede	0	0	0	0
abde	200	52	0.1688	0.3512
abcde	0	0	0	0

6 Combining Evidence.

Suppose cow *c1* got treatments *b, c, d, e..* We have no data about such cows. But she also got all but one of the four treatments — i.e., *bed, bce,* etc. In particular, she is *cde* and also *bee.* But these two possible reference classes disagree in the sense that neither simply overlaps the other. Put otherwise, they disagree in the sense that there is a frequency or chance that is compatible with one and not compatible with the other. The frequency in *cde* is in [.2423,3577], and the frequency in *bee* is in [.1803,-2448]. Nothing seems to indicate any way of reconciling or combining these two possible reference classes — in particular, we have no data on their intersection. Should we give up and say that the interval [0,1] characterizes the uncertainty to be attached to this cow?

There are two natural alternatives. One corresponds to the Dempster/Shafer updating procedure [Shafer, 1976]. This is to look at the set of pairs $<x,y>$ such that *x* belongs to *cde* and *y* belongs to *bce.* Now the cow in question corresponds also to a pair of the form $<z,z>$ We don't know the frequency of success in the diagonal, but we do know one thing about the pair $<c_1,c_1>$: that it either belongs to the set of pairs both of which founder, or to the set of pairs both of which do not founder. This leads us to a subset of the set of pairs in general — namely those that agree with respect to success. And we can calculate bounds on the relative frequency of success in this subset: they are given exactly in accordance with the Dempster/Shafer formula for combining evidence: $p_1p_2/(1-p_1-p_2+2p_1p_2)$. This construction for combining evidence has been suggested by Ron Loui [1986].

But faced with this concrete example, it does not seem so plausible.[1] Suppose that we know that *a* is in R_1 and also in R_2, and that '$\%(T,R_1) \in [.3,.4]$ and $\%(T,R_2) \in [.40,.45]$. These intervals conflict. But consider $<a,a>$ as a member of $R_1 \times R_2$. It must belong to that part of $R_1 \times R_2$ that is included in the union of T x T and $\sim T \times \sim T$. It belongs to $T \times T$ just in case *a* belongs to T; and the frequency of this kind of occurrence is in $[.22,.35]$, by the preceding formula.[2] In the particular case at hand, the resulting interval is [.0844,.1520]. But this seems counterintuitive, since it is less than the frequency in either of the two classes whose information we have combined.

An alternative approach would be to argue thus: We find ourselves in a conflict situation because we have evidence that supports two intervals that disagree. But the same evidence supports, *even more strongly,* two intervals that do not disagree. Namely: the intervals bounded by the lower of the two lower bounds, and the higher of the two upper bounds.

The proportion of foundered cows in the class *bce* is known to be between .1803 and .2466, and therefore obviously between .1800 and .3577 And the proportion of foundered cows in the class *cde* is known to be between .2423 and .3577, and therefore also between .1803 and .3577. These limits seem intuitively to be reasonable limits for *c1* — the probability that she will founder is [.1803,3577]. This is a broad limit; one that in the face of conflicting evidence does not seem unreasonable.

How about the other classes that *c1* is known to belong to: *bc, bd,* etc.? Each of them conflicts with one or another of the more specific classes; but such a conflict is most reasonably resolved by ignoring the less specific class.

7 Precision

Contrast cow C2- She is known to be an *abe;* therefore she is also known to be an *a* a *b* an *e,* an *ab,* a *be,* and an *ae.* These items of statistical knowledge yield:

abc	.2046	.2681
ab	.2235	.2591
ac	.2225	.2661

[1] While we don't know the relevant frequency among instances of the form $<x,x>$, in general we allow ourselves to use frequencies known to apply to broader classes as the basis for the probability applicable to an instance in a narrower class, else we would be stuck with the [0,1] characteristic of a unit class.

[2] It might be thought that the process could be iterated, and that it would lead to intervals that migrate toward the extremes of 1 and 0. This is not so, since the "higher dimensionality" constraint in the full system prevents the components from "disagreeing" with the cross product subset [Kyburg, 1974,1983,1990]. Nevertheless, as Bulent Murtzaoglu [1990] has pointed out, the results are counterintuitive even without iteration.

bc	.2217	.2761
a	.2374	.2670
b	.2389	.2711
c	.2366	.2738
u	.2493	.2751

as the intervals in which the frequency of founder might lie. Although *abc* is the "smallest" class in which *C2* is known to be, our knowledge about that class does not conflict with knowledge about any of the superclasses in which it is known to be included; it is merely less precise. If we look at the superclasses, the one about which we have the most information is *ab;* this does not conflict with any of the others, and thus we use *ab* as the reference class and take the probability to be 1.2235,.2591]. [3]

Consider *c3* who is known to belong to *ade,* in which the relative frequency is in [.1688,3512]. The corresponding table is:

ade	.1688	.3512
ad	.2150	.2666
ae	.2070	.2470
de	.2311	.3056
a	.2374	.2670
d	.2368	.2789
e	.2217	.2568
u	.2493	.2751

Here, although *ad* contains more precise information, and does not conflict with *ade,* it does conflict with *ae* and *de,* who have an equal right to be heard. Since the subsets *ad, ae,* and *de* also conflict with each other, and none of them conflicts with our knowledge about *ade,* neither does their cover conflict with our knowledge about *ade.* All three of these classes are known to exhibit a frequency of founder in the interval [.2070,3056]. This is more precise knowledge, and should be used to determine the probability.

Finally, suppose *C4* is known to belong to *abd* The table is:

abd	.2109	.2672
ab	.2235	.2591
ad	.2150	.2666
bd	.2240	.2778
a	.2374	.2670
b	.2389	.2711
	.2368	.2789
u	.2493	.2751

Here, although again more is known about *ab* than about *abd,* and there is no conflict, the superset *bd* conflicts with the more specific set (and therefore also with *ab),* and so we take the probability to be determined by our knowledge about *abd:* [.2109,-2672].

The principle we have been following so far is to reject a candidate reference class whenever the relevant frequency in it, or in any other candidate reference class of equal specificity, differs from that in the most specific candidate reference class, where "differs" means that neither interval is included in the other. The result will be a set of intervals, none of which differ from that of the most specific class. We use the most informative interval for our probability.[4]

Note that we are taking the probability to *be* the interval, not that it is "included in" the interval. There may be no identifiable class that contains a more precisely specified proportion than the one we are working with.

8 Collections of Herds: Bayesian Inference.

Suppose that we have obtained data on the relation between the five factors we are considering and founder, not merely from one herd, but from several. The following table gives relevant data concerning four herds. The herds quite clearly differ in the effect of the various factors. This might be due to genetic differences among the breeding lines that dominate the various herds, or it might be due to environmental factors. The upper and lower frequencies, as before, reflect conservative confidence limits at the 0.90 confidence level.

	herd 1		herd 2	
	10000		6000	
	fl	fu	fl	fa
u	0.249	0.275	0.408	0.442
b	0.239	0.271	0.434	0.478
c	0.237	0.274	0.427	0.47
ab	0.224	0.259	0.453	0.503
bc	0.222	0.276	0.479	0.535

	herd 3		herd 4		all	
	3000		20000		39000	
	fl	fu	fl	fu	fl	fu
u	031	035	0.304	032	0312	0.33
b	0.32	0.38	0.262	029	0.294	031
c	*029*	035	0.286	031	0.306	032

[3] If there was conflict between *ab* and another superclass two cases could arise. First, there is conflict also with the original subclass. Then that subclass would prevail. Second, neither of the superclasses conflicts with the subclass. Then the conflict between the superclasses can be resolved by a cover, and since the cover will not conflict with the subclass either, it can plausibly be taken to determine the probability.

[4] The "most informative" interval may be an interval cover, as we have seen.

ab	0.38	0.46	0202	024	0.278	0.3
be	0.3	038	0.221	026	0.29	0.31

What is the chance of cow *c$* , who belongs to one of these herds and has received treatments *b* and *c* coming down with founder? You may ask, "Which herd does she belong to?" but we are assuming that we don't know: all we know is that she belongs to one of these four herds. Like the preceding case, in which we did not know which treatments a cow did *not* get, we are here dealing with a situation of limited knowledge.

The candidate classes are *u, b, c,* and *bc:*

b	.2935	.3105
c	.3061	.3239
bc	.2900	.3144
u	.3122	.3253

The intuitively most plausible reference class is *bc,* even though our knowledge about *be* is much less precise than our knowledge about *b* and about *c.*

Now let us suppose that cow c6 is selected from these herds by an agricultural agent who first chooses a herd at random to sample from, and then chooses a cow from the herd he has chosen.

We have a new variable to take account of: herd number. The herd is chosen at random, which is to say that in the long run, each herd will be selected with equal frequency — i.e., 1/4. We know how to calculate the probability: it is the average of the frequencies in the herds. The probability that c6 suffers from founder is [3057,-3608].

What is the reference class that gives rise to this interval? It is perfectly straight-forward: it is the set of pairs <x,y> such that *x* is a herd, and *y* is a cow from that herd; the target set, of course, is the set of such pairs of which the second member founders. Let us call the reference set just described B, and the target set S.

While it is easy to agree to that probability interval, it is less easy to see why the former answer is wrong. The cow c6 is, now as before, a member of the total collection of cows about which we have data. What prevents that from being a correct reference class, as it was in the previous case?

The principle required — from which the specificity principle follows as a special case — is what we shall call the Bayesian Principle. Suppose you have a reference set of the form described, and that it leads to an interval that conflicts with an interval obtained from some other reference set for the same proposition, for example, the reference set consisting of all the cows in all four herds. We can get from this reference set to the correct one as follows: Note that the correct reference set makes use of a set of predicates (forming a partition) that does not appear in the first reference set. Form the cross product of the partition and the original set. (This is the set of all pairs consisting of a herd and a cow from the collection of four herds.) Then note that the intuitively obvious reference set is a subset of that set of pairs — the subset in which the cow (second member of the pair) is a member of the herd (the first member of the pair).

This principle implies the specificity principle because if we use a vacuous partition, we will obtain the simple subset rule.

In general, when a Bayesian principle is appropriate, it is because it is appropriate for the reference set to be construed as resulting from a two (or multi) stage process. What can defend the probability that results from that process is precisely the construction just described.

9 Sizes of Samples: Statistical Inference

Suppose we want to know how many cows in the class *ab* will founder in the long run. We have a lot of evidence. In particular, we have the evidence obtained from herd 1: of 5250 cows known to belong to *ab,* 1267 came down with founder. It follows, at a confidence level of .90, that between .2235 and .2591 of cows belonging to *ab* will founder.

We also have the evidence obtained from all four herds: of 14,800 animals having those treatments, 4272 foundered, yielding an observed frequency of .2886, and a 90% confidence interval of *1.2780,29931*

The first inference gives us a perfectly proper inference for inferring the proportion of cows falling in the class *ab* who founder. If that was the only herd we had investigated, we would indeed be entitled to take that as a general conclusion about cows in general. In the presence of the broader sample, of course, it is no longer appropriate. Why not?

Because it does not take into account all the evidence we have about cows and foundering. If what we are interested in is the proportion of cows in general that founder among those who fall into the class in question, then we should take into account all we know of cows falling into that class.

But that may not be what we are interested in. Though we have more information about cows in general than we have about cows from particular herds, it may be that we always know what herd a cow comes from. If that's the case, it will be readily seen from what we have said before that what is of interest to us will be the proportion of cows that founder within the smaller class: *ab and in herd 1*; *ab and in herd* 2, etc. Thus the first inference may be perfectly all right, so long as the class about which we are making the inference is *ab and herd* 1 rather than *ab* in general.

In short, what we need to have justified are all the inferences we can legitimately make. Which ones we should make may be a function of our particular situation.

10 Conclusions*

If we identify probability with empirical frequency, the probabilities of specific eventualities are either unknown (or are 0 or 1) since those are the only frequencies that can occur in a single instance. The frequency approach can be construed less literally, but then we must devise a way of combining all the evidence we have bearing on a specific eventuality. Frequency approaches to probability provide no machinery for combining evidence, or for choosing among reference classes.

If we identify probability with mere subjective opinion the connection with empirical data is equally unanalyzed. Conditionalization is a consistent way of updating opinions, but since our opinions are uncertain to start with, the results of conditionalization can be taken to cast doubt on what we took to be our initial opinions.

The evidential approach to probability attempts to formalize the evidential relation between general knowledge of frequencies and specific individual events. To do this requires first that the description of the individual event be related to a class about which we have statistical knowledge. In general there will be more than one way of doing this.

Second, we need some way of adjudicating conflicts, if any, among the statistics corresponding to the various ways of relating the event in question, under any acceptable description, to various possible reference classes. We considered three kinds of conflict: theser are discussed in more detail in [Kyburg, 1983] and in [Kyburg, 1974].

Third, there are two kinds of conflicting statistics: Since in general our statistical knowledge related to an individual event will be in the form of intervals, we will encounter cases in which one interval is included in another. If there is no other conflict, it makes sense to use the tightest interval we can justify. But it is also possible that the two intervals will be disjoint, or overlap without inclusion. This is the difficult case.

Evidential probability provides machinery for dealing with the individual instance that concerns us, employing all the statistical evidence we have bearing on that instance. It thus serves where subjective views would use mere opinion. It also takes every probability to be based on general statistical knowledge, thus satisfying one of the intuitive desiderata for which the frequency theories were designed. And in the world of expert knowledge, we may often (always?) assume that the opinions of the expert are, or should have been, or could have been, based ultimately on knowledge of frequencies.

[Cheeseman, 1988] Peter Cheeseman. Inquiry into Computer Understanding. *Computational Intelligence* 4,: 58-66, 1988.

[de Finetti, 1937] Bruno de Finett. La Prevision: Ses Lois Logiques, Ses Sources Sujectives. *Annales De L'Institute Henri Poincare,* 1937.

[Kyburg, 1961] Henry E. Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, Connecticut, 1961.

[Kyburg, 1974] Henry E. Kyburg, Jr. *The Logical Foundations of Statistical Inference.* Reidel, Dordrecht, 1974.

[Kyburg,1983] Henry E. Kyburg, Jr. The Reference Class. *Philosophy of Science* 50, pp. 374-397, 1983.

[Kyburg, 1990)] Henry E. Kyburg,Jr. Uncertainty and the Conditioning of Beliefs. In George M. von Furstenberg,ed. *Acting Under Uncertainty: Multidisciplinary Conceptions,* Kluwer, Dordrecht, pages 77-94, 1990.

[Loui, 1986] Ronald P. Loui. Computing Reference Classes *Proceedings of the 1986 Workshop on Uncertainty in Artificial Intelligence,* pages 183-88, 1986.

[McCarthy and Hayes, 1969] John McCarthy and Pat Hayes. Some Philosophical Problems From the Standpoint of Artifical Intelligence. In Meltzer and Michie (eds), *Machine Intelligence* 4, Edinburgh, pp. 463-502.

[Murtezaoglu, 1990] Biilent Murtezaoglu. A Modification to Evidential Probability. TR, University of Rochester, 1990.

[Raiffa and Schlaifer 1961] Howard Raiffa and Robert Schlaifer. *Applied Statistical Decision Theory,* MIT Press, Cambridge, Mass, 1961.

[Ramsey, 1931] F. P. Ramsey.The *Foundations of Mathematics and Other Essays,* Humanities Press, New York, 1931.

[Reichenbach, 1949] Hans Reichenbach. *The Theory of Probability* University of California Press, Berkeley and Los Angles, 1949.

[Savage,1954 L. J. Savage. *Foundations of Statistics,* John Wiley, New York, 1954.

[Shafer, 1976] Glenn Shafer. *A Mathematical Theory of Evidence,* Princeton University Press, Princeton, 1976.

[Wilson, 1952] E. Bright Wilson, Jr.. *An Introduction to Scientific Research,* McGraw-Hill, New York, 1952.