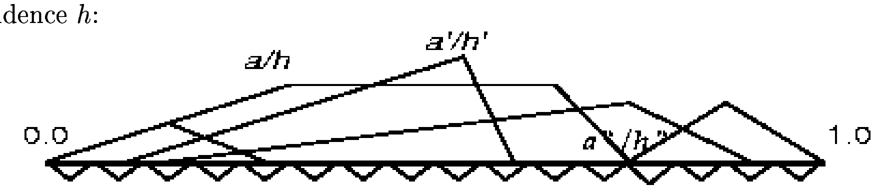
INTERVAL-VALUED PROBABILITIES

HENRY E. KYBURG, JR.
UNIVERSITY OF ROCHESTER
KYBURG@CS.ROCHESTER.EDU

This is a contribution to the *Documentation Section* on the website of the *Society for Imprecise Probability Theory and Applications* (SIPTA): `http://www.sipta.org`

1. Background.

It was John Maynard Keynes [Key52] who first forcibly argued that probabilities cannot be simply ordered. There are cases, he argued, in which the probability of the hypothesis h can be regarded neither as greater than that of hypothesis k, nor less than that of hypothesis k, nor yet equal to that of hypothesis k. Although he did not provide a mathematical structure for his probability values, he did give us some hints: Every probability is comparable to the 0 probability and to the 1 probability. In general, some probabilities (those that can be based on a correct application of the principle of indifference) have rational numerical values, and can serve to bound those that are not precisely comparable to rational valued probabilities. The structure looks something like this, where a/h represents the (not necessarily numerical) probability of the proposition a relative to the (total) evidence h:



The points on the bold line represent the strength of arguments that can be measured by real numbers; the other points represent other probability values; the edges represent the possibility of comparison. There is, for example, no relation ($<$, $>$, $=$) between a/h and a'/h' in the diagram. The sawtooth line reflects the fact that even when the principle of indifference can be applied, there may be arguments whose strength can be bounded no more precisely than by an adjacent pair of indifference arguments. Note that a/h in the diagram is bounded numerically only by 0.0 and the strength of a''/h''.

Keynes' ideas were taken up by B. O. Koopman [Koo40a, Koo40b, Koo41], who provided an axiomatization for Keynes' probability values. The axioms are qualitative, and reflect what Keynes said about probability judgment. (It should be remembered that for Keynes probability judgment was intended to be objective in the sense that logic is objective. Although different people may accept different premises, whether or not a conclusion follows logically from a *given* set of premises

Date: September 1, 1998.

Support for this work was provided by National Science Foundation grant IRI-9411267.

is objective. Though Ramsey [Ram22] attacked this aspect of Keynes' theory, it can be argued that Keynes never completely gave in.)

Koopman provided qualitative axioms for probability judgment that yielded an algebra of probabilities. This algebra included a set of rational-valued numerical probabilities, based on the idea that we could always (hypothetically) construct a deck of n cards such that the probability of getting m specified cards would be the ratio m/n. Let us call these probabilities rational-valued probabilities. An arbitrary probability value from the algebra may then be represented by the greatest lower bound of the rational probabilities it is judged greater than, and the least upper bound of the rational probabilities it is judged less than.

Terrence Fine, in a well known work *Theories of Probability* [Fin73], provides a careful analysis of probability developed step by step from qualitative, to comparative probability relations, to quantitative probability functions. He examines the claims of Keynes and the structures developed by Koopman, but very early on in this development — even for comparative probability — assumes that the probability relation is complete: for any two propositions, either one is more probable than the other or they are equally probable. This does not hold for Keynes' probabilities, nor for probability as characterized by the axioms of Koopman. Indeed, it precludes interval valued probabilities, which have a certain amount of intrinsic appeal, and which seem to conform to Keynes' intuitions [Kyb97]. Once one has the completeness of the probability relation, it is only a short hop and a skip to real-valued probabilities. In later work, for example in [Fin88], interval-valued probabilities are strongly advocated by Fine.

2. Motivations

There are two motivations for turning to interval valued probabilities. One, which may be attributed to C.A.B. Smith [Smi61], and is found in its most highly developed state in Walley [Wal91], is to describe the states of individuals with regard to their degrees of belief more adequately and more realistically than standard subjective Bayesianism can. The classical view of personalism, as proposed by L J. Savage [Sav54], assumes that we can be forced to specify our degree of belief in a proposition to any number of decimal places. This is not only unrealistic, but verges on inconsistency, since the procedures assume that the payoffs do not reflect the value of the sanction that makes choice necessary. Thus I am not *allowed* to say "I don't know," or to choose "none of the above," or to choose in an undetermined, arbitrary, or whimsical way, when faced with the choice of a ticket that returns a dollar if Q. Albert is elected, and a ticket in a fair million-ticket lottery that returns a dollar if one of 579,321 specified tickets is chosen. "Not allowed" can only mean that there is some sanction, whose disutility I wish to avoid, that attends my failure to answer the question. But of course that disutility should be taken account of in the behavioral assessment of my degrees of belief, and the fine gradation of degrees of belief may be lost.

The other motivation reflects the idea that our degrees of belief ought to be objectively determined, if they are to count as rational, and thus should reflect our statistical knowledge; since our statistical knowledge is never precise, our beliefs are never constrained more precisely than by sets of statistical distributions. This was the motivation underlying the development in *Probability and the Logic of Rational Belief* [Kyb61] and more recently [Kyb74] and [Kyb97]. We shall examine the two kinds of motivation separately, the first in sections 3 – 6; the second in section 7.

3. Probability Intervals as Descriptive

I. J. Good [Goo62] takes the necessity for intervals to arise from the fact that our initial qualitative judgments are imperfect. If we were perfect logicians, and had perfect self knowledge, we might be perfect point-valued Bayesians, but in fact the judgments that are input to our scientific black box are only partially ordered. (p. 322). Good obtains upper and lower probabilities as constraints on the numerical output (discriminations) produced by qualitative inputs (judgments of the form $P(E|H) \leq P(F|G)$). From the properties of point-valued probabilities he derives (quite simply) a set of axioms similar to those provided directly by Koopman. It is the nature of human judgment that leads to the intervals that are the output of the 'black box.'

It is true also of the ideas developed by C. A. B. Smith [Smi61, Smi65] that the data we can obtain on degrees of belief only roughly constrain the (true?) probabilities. To offer to bet at odds of m to n on the truth of a proposition S, is to be willing to put up m units of utility, which will be lost if S turns out to be false, in return for the chance of winning n units of utility if S turns out to be true. These odds correspond to the assignment of a probability of $m/(m+n)$ to S. The agent, according to Smith, who is willing to bet at odds of m to n will also be willing to bet at odds of $m-k$ to n. Smith defines the 'lower pignic odds' [Smi65, p. 5] to be the upper bound on the odds that the agent is willing to offer in a bet on an event. 'Upper pignic odds' are similarly defined.

These odds lead to lower and upper probabilities by means of the transformation just mentioned. The development of the 1961 paper focuses mainly on the existence of (and constraints on) 'medial' probabilities – i.e., probabilities that strictly fall between the upper and lower probabilities. Thus, for example, if we have exclusive events A and B, it follows from the natural constraints on upper and lower pignistic odds that there exist medial probabilities pA, pB, and p, lying between the upper and lower probabilities of A, B, and $A \cup B$, respectively, such that $p = pA + pB$.

The representation of uncertainty can thus be looked on as a representation employing sets of probability functions: upper and lower probabilities are bounds on sets of medial probabilities. Strictly speaking, of course, these upper and lower 'probabilities' are envelopes of probabilities rather than being probabilities themselves: for obvious reasons, they do not satisfy the standard probability axioms.

It is perfectly straight-forward to consider the updating of sets of probability functions by conditionalization. The conditional probability of A *given* B, $P(A|B)$, is simply the ratio of $P(A \cap B)$ to $P(B)$, for each probability function P: The set of values of the conditional probability $P(\,.\,|B)$ is determined by the set of original probability functions. If the original set is convex, so will the set of conditional probabilities be convex. There is the question of what to do about conditioning on events of 0 probability. The simplest approach is simply to disregard the set of probability functions for which $P(B) = 0$. An alternative is to take probability to be a two place function in the first place (so that conditional probabilities are always defined [Pop57]) and to derive probability intervals from sets of these two-place functions.

4. A Behavioral Interpretation of Intervals

A nearly encyclopedic analysis of an approach that takes intervals to be basic is provided by Peter Walley [Wal91]. (For a brief characterization, see [Wal96] Walley starts from a behavioral view much like that of Smith, but rejects the assumption that the upper and lower probabilities are to be construed as envelopes of probability functions. Standards of coherence are imposed directly on lower probabilities: a lower probability P is coherent if it does not lead to a sure loss, and if it embodies the transitivity of preference for gambles (p. 29). It is an important (and non-trivial) theorem of this approach that for the coherent agent upper and lower probabilities may be represented as envelopes of sets of probability functions. In Walley's notation [Wal91, p. 134], where $\underline{P}$ is a coherent lower probability function:

(1) $\underline{P}$ avoids sure loss if and only if it is dominated by a classical probability; and

(2) $\underline{P}$ is coherent if and only if it is the lower envelope of some class of classical probabilities.

It is important to be clear about what is going on here. The fact that upper and lower probabilities can be construed as sets of distributions does not mean that they *must* be so construed. The statistical theory that emerges, for one thing, is quite different. Furthermore, definitions of various structural properties, such as independence and permutability can be given directly in terms of these lower probabilities, and these definitions do not correspond directly to the classical concepts applicable to classical probability functions. For example, according to Walley's view, two events are independent when betting rates on either one, conditional on the other, are the same as the unconditional betting rates; according to the conception that takes the probability functions of which the lower probability is an envelope, two events are independent if they are independent according to every probability within the envelope. These are clearly two quite different things.

Walley makes much of the fact that he can give direct behavioral interpretations to these concepts, as well as to lower probabilities. Since the behavior of people rarely conforms to any rational standards — including those that Walley takes as axioms for lower probabilities — it is not clear how much weight should be given to this aspect of his theory. On the other hand, that behavioral norms can be derived independently of the existence of any underlying probability functions seems to be a telling point.

Walley is concerned primarily with the foundations of statistical inference, but also explores in detail questions concerning the representation of beliefs or judgments. In Chapter 4 he discusses a number of ways in which the assessment of beliefs can yield an interval characterization. Given a space Ω the agent may make any finite number of judgments, including classificatory judgments ('A is probable'), comparative judgments ('A is more probable than B'), ratio judgments ('C is at least twice as probable as B'), as well as direct judgments ('the probability of D is at least 0.3'). These judgments may also include the use of neighborhood models, and upper and lower distribution functions. In each case the problem discussed is primarily that of finding an appropriate interval representation based on a feasible amount of inquiry.

Note that the special case in which the agent is to make precise numerical judgments on the atoms of the possibility space Ω may require that the agent solve an exponential problem: if the language of Ω contains n atomic formulas, Ω will contain 2^n elements, and require that many judgments on the part of an agent who wants a point-valued probability. Without simplifications (for example a principle of indifference) this is an intractable problem.

Thus there is indeed a virtue in the interval representation from a purely descriptive point of view. Given a finite, feasible, amount of inquiry, we can determine the beliefs of an ideally rational coherent agent only up to a lower probability, which yields an upper probability through the identity $\overline{P}(H) = 1 - \underline{P}(\neg H)$, and thus an interval on each proposition. (Note that this procedure does not yield a single classical probability function.) This is theoretically and practically important because there do not need to be many propositions involved before 2^n — the number of state descriptions we can generate from n logically independent atomic statements — becomes unwieldy, and of course there is no way of soliciting the infinite number of judgments required to yield a continuous distribution.

Given a finite number of judgments, there are two remaining problems: first, to make sure that those judgments avoid sure loss i.e., that there exists a classical probability function that satisfies the constraints embodied in those judgments, and second, to extend those judgments to include judgments derived from them. This leads to the closure under finite mixtures of the gambles corresponding to those initial judgments. Thus we can obtain lower probability functions by means of a feasible finite set of judgments.

One of the most attractive features of Walley's treatment of interval-valued probabilities is that it naturally lends itself to sequential use. Constraints on the lower probability that may be taken to represent the agent's opinions may be introduced one at a time, checked for consistency, and used to modify the existing lower probability, until the representation is deemed adequate to the problem at hand. Furthermore, the constraints themselves need not be taken as perfectly precise: that an agent will accept a bet on A at less than even money may indicate a constraint to the effect that $\underline{P}(A) = 0.5$; it does not rule out the possibility of refining that constraint at a subsequent stage to $\underline{P}(A) = 0.50342$. The usual personalist technique of requiring that the agent pick a number such that he would be willing to take either side of a bet at odds determined by that number precludes the possibility of this kind of refinement.

The updating of probability intervals in Walley's system is non-trivial. First, we must take account of the fact that we cannot simply look at the underlying probability functions and perform classical conditionalization: there may be no such functions! Second, it can easily happen that that the lower probability of the event on which we want to condition is 0.0. If we look at the intervals as derived from classical probability functions there may be some justification for ignoring those that assign probability 0 to the conditioning event; on Walley's view there is no such justification. Chapter Six of [Wal91] contains an extensive analysis of the complexities involved.

5. Belief functions

A representation of belief that is presented as an alternative to the classical personalistic representation is provided Shafer [Sha76]. The theory of belief functions assigns a number, $Bel(A)$, to each subset A of a frame of discernment (universe of possibilities) Ω. The function Bel satisfies the axioms

(1) $Bel(\emptyset) = 0,$

(2) $Bel(\Omega) = 1,$

(3) $Bel(A_1 \cup \cdots \cup A_n) = \sum_{\emptyset \neq \mathrm{I} \subset \{1,\ldots,n\}} (-1)^{|\mathrm{I}|+1} Bel(\bigcap_{i \in \mathrm{I}} A_i).$

It has been shown [Dem68, Kyb87] that to every representation of beliefs in terms of belief functions, there corresponds a representation in terms of sets of probabilities. This set of probability functions will be convex: given any two functions in the set, their $a : (1 - a)$ mixture will also be in the set. When it comes to the representation of degrees of belief, this makes perfectly good sense. It is difficult to imagine grounds on which an agent would be inclined to regard 2:3 odds on as fair (correspondiong to a lower probability of 0.4), and also inclined to regard 2:3 odds against as fair (corresponding to an upper probability of 0.6), but would not regard even money as fair. As we shall see later, construing set valued probabilities as reflections of objective statistical reality may put a different light on the issue of convexity.

It has also been shown [Kyb87] that there are convex sets of probabilities that do not have a representation as belief functions. It could thus be argued that belief functions correspond to a special case of the convex set representation.

While this is true formally, it can be claimed that the interpretation of belief functions is quite different from that of convex sets of probabilities. A belief function represents "the impact of evidence" rather than a "behavioral disposition", as might a set of probabilities. In itself it is not clear what this distinction comes to: perhaps it is that all the evidence I have could support the proposition A to the degree 0.44 ($Bel(A) = 0.44$), and yet I might not be willing to pay more than \$.40 for a ticket returning a dollar on A. This is just to say that my behavioral dispositions could be less precise than the evidence requires. Nevertheless, in accordance with Walley's model, we could think of a belief function as simply imposing another constraint on the intervals that are being taken to represent our distribution of belief.

When it comes to updating, however, matters are quite different. Belief functions come with their own rule of combination, "Dempster's rule of combination," [Sha76, Dem68] which does not correspond to conditionalization. Suppose Bel' represents a belief function that corresponds to the acceptance of evidence E. Then the rule of combination leads to the simple updating formulas,

$$Bel'(E) = Bel(H|E) = \frac{Bel(H \vee \neg E) - Bel(\neg E)}{1 - Bel(\neg E)},$$

for belief, and

$$P^{*\prime}(E) = P^*(H|E) = \frac{P^*(H \wedge E)}{P^*(E)}$$

for plausibility P^*, the upper bound on credence, where

$$P^*(A) = 1 - Bel(\neg A).$$

It is shown in [Kyb87] that belief functions combine in a way that yields new belief functions that are representable by a proper subset of the corresponding set of conditional probabilities. For example, let H be a hypothesis and E be some evidence concerning it. Suppose that we have the following lower probabilities (which also determine a belief function, though not all sets of lower probabilities do):[1]

$\underline{P}(H) = 0.5$	$\underline{P}(\neg H) = 0.4$
$\underline{P}(E) = 0.4\underline{P}(\neg E) = 0.3$	
$\underline{P}(HE) = 0.2$	$\underline{P}(H\neg E) = 0.2$
$\underline{P}(\neg HE) = 0.2$	$\underline{P}(\neg H\neg E) = 0.1$
$\underline{P}(HE \vee H\neg E) = 0.5$	$\underline{P}(H\neg E \vee \neg HE) = 0.5$
$\underline{P}(HE \vee \neg HE) = 0.4$	$\underline{P}(H\neg E \vee \neg H\neg E) = 0.3$
$\underline{P}(HE \vee \neg H\neg E) = 0.3$	$\underline{P}(\neg HE \vee \neg H\neg E) = 0.4$
$\underline{P}(HE \vee H\neg E \vee \neg HE) = 0.8$	$\underline{P}(HE \vee H\neg E \vee \neg H\neg E) = 0.6$
$\underline{P}(HE \vee \neg HE \vee \neg H\neg E) = 0.6$	$\underline{P}(\neg HE \vee H\neg E \vee \neg H\neg E) = 0.6$

Let us now update the belief in H *given* the evidence E. It will be easily verified that $\underline{P}(H|E) = 4/14$; $Bel(H|E) = 6/14$; $P^*(H|E) = 7/14$; and $\overline{P}(H|E) = 8/14$.

This could, of course, be regarded as a virtue of belief functions: their rule of combination gives more precise results than the corresponding application of conditional probability. This is not an uncontroversial advantage. It could be argued

[1] These lower probabilities correspond to the following mass function: $m(HE) = m(H\neg E) = m(\neg HE)$; $m(\neg HnegE) = m(HE \vee H\neg E) = m(\neg HE \vee H\neg E) = m(\neg HE \vee \neg H\neg E) = 0.1$.

that this precision is fraught with peril, at least if the probabilities in the corresponding envelope are in any sense "objective," for we could then be arguing from objective probabilities to (uncertain) conclusions that represent probabilities that transcend our evidence. It is perhaps these considerations that lend the greatest support to the claim that belief functions have nothing to do with probabilities. They do not combine as probabilities do.

Nevertheless, belief functions must be mentioned here in view of the fact that, barring the question of updating, they can be represented as (convex) sets of probability functions or as probability envelopes.

6. Transferrable Belief

Although the representation by belief functions formally corresponds to a representation in terms of convex sets of probability, there can be a significant difference in interpretation. This difference is exploited in the transferable belief model of Smets [Sme88, SK94]. According to this model, the *credibility* of a proposition A is represented directly by a number $Cr(A)$; this number is determined not as the lower bound of a set of numbers corresponding to a set of probability functions, but through the "basic probability assignment" underlying belief. This basic probability assignment assigns a measurs m to the subsets of the set of possible worlds, just as for Shafer. There is no direct connection between credibility and behavior (as distinct from Walley's view), and no direct connection between credibility and probability. This distinction of interpretation has two important consequences.

First, it allows a variety of procedures for combining evidence. We have just observed that the rule of combination that goes with belief functions is controversial. If we construe belief functions as lower envelopes of probability functions, it is not only controversial, but perhaps even wrong: there are arguments that probabilities must be combined by conditionalization, and conditionalization, applied to convex sets of probability functions may conflict with the result of applying the rule of combination. To the extent that these arguments are right (they themselves are not uncontroversial!) the rule of combination, under the set of probabilities interpretation of belief functions, is wrong.

Under Smets' transferable belief model nothing is "built in" concerning the updating procedure. There are several procedures, including conditionalization and Dempster conditioning, as well as imaging [Lew76], and others, for modifying belief distributions in the light of new evidence. Which one or ones are appropriate under what circumstances is a matter for investigation, on Smets' view. This is a great advance over the rigid adherence to Dempster's rule required by Shafer in 1976.

Second, the transferable belief model allows us to focus on a single probability function for a given practical framework — a betting frame. While the convex set of probabilities treatment — for example, that of C. A. B. Smith [Smi61, Smi65] — treats all members of that convex set equally, the transferable belief model provides a mechanism for selecting a particular probability distribution (a distribution of pignistic probabilities determining betting odds) corresponding to a credibility function Cr and a betting frame, by means of a special version of the (historically notorious) principle of indifference. The fact that the betting frame is involved allows for the possibility two propositions, A and B, known to be equivalent, and therefore having the same credibility value, may have different pignistic probabilities.

7. Statistical Knowledge

Our treatment of uncertainty may be taken to depend on our having objective statistical knowledge. If so, then we must ask about the source of that knowledge.

The source is generally a matter of statistical inference, and so the foundations of statistical inference is where we must look for enlightenment concerning statistical knowledge. Alas, we will find the foundations of statistical inference to be as controversial as the representation of uncertainty in the first place. In fact, many of the same issues are involved. While one approach to statistical inference demands that it depend on prior probability distributions, and requires that the output of statistical inference be a distribution over a family of statistical hypotheses, another approach to statistical inference demands that it depend on objective frequency distributions, and that its output be the rejection of a set of statistical hypotheses. (Almost no one talks of "accepting" statistical hypotheses, but it is widely regarded as All Right to fail-to-reject a collection of hypotheses.)

A promising line of attack, explored by [Wal95] in one special case, takes the set-of-distributions approach to statistical inference itself. For example, if we have a population that we can characterize as binomial $B_{m,p}$ for p in [0,1], then, whatever p may be, the distribution of a relative frequency close to that in the parent population itself among all m-membered samples of that population can be determined by making use of the Dirichlet distribution. There is thus the possibility of an objective, or largely objective, method of statistical inference that yields intervals, corresponding to sets of statistical hypotheses.

The next controversial issue has to do with whether the result of such a statistical inference is the assignment of a probability to a set of statistical hypotheses, or the acceptance of those hypotheses. Do we infer "The degree of certainty of the interval valued hypothesis that the process is binomial with a parameter $p \in [p_u, p_m]$, relative to what we know, is 0.95," or do we infer "The process is binomial with a parameter $p \in [p_u, p_m]$" (and the evidence we have allows us to infer this with acceptance level, or confidence, 0.95).

This distinction was called forcibly to our attention by Carl Hempel [Hem61], and is still a matter of controversy. It is, in fact, at the heart of the controversy between "probabilists" who take the output of uncertain inference to include an index of the uncertainty, and "logicists" who argue that nonmonotonic inference issues in categorical statements (like the second one quoted above) that are nevertheless subject to withdrawal in the light of new evidence. For further discussion of this distinction, see [Kyb88].

This is not the place to review this controversy, though it is interesting and relevant that a matter so basic has become an issue in such a wide variety of contexts. In any event, if we assume that statistical inference can yield the acceptance of anything, it must the the acceptance of sets of hypotheses, such as "The process is binomial with a parameter $p \in [p_u, p_m]$." If we want to ground uncertainty in statistical knowledge, it must be *approximate* knowledge. This is the approach followed in [Kyb74] and [Kyb97].

It follows that since the statistical knowledge we obtain inductively can only be approximiate, uncertainy itself can at most be measured by intervals. This is so quite independently of the propensity of the agent to accept or reject bets, quite independently of the agent's belief function (hypothesized by Shafer), quite independently of the agent's credibility function (Smets). If our uncertainties are determined objectively, by what we know of the empirical world, they must generally be represented by intervals.

8. Reconciliation

Subjective probability doesn't take itself too literally; it either has an explicitly normative element, or it is construed as the theory of belief of a "rational" individual. The individual whose lower belief in A is 0.4, and whose lower belief in $\neg A$ is

0.7 is, by anybody's standards, irrational. Thus the question is not whether or not a plausible subjective view embodies constraints other than that of reflecting how people really order their beliefs, but the degree and source of such constraints.

Ramsey's view was that the axioms of the probability calculus functioned like laws of logic: they represented, on his view, all the constraints that you could impose on the credibilities of a rational individual. "... we do not regard it as belonging to formal logic to say what should be a man's expectation of drawing a white or a black ball from an urn; his original expectations may within the limits of consistency be any he likes..." [Ram31, p. 189]. Carnap and other writers thought that some more constraints could be imposed. These are constraints expressed in terms of symmetries [Car50, Car80] or entropy [Jay58]. Similar constraints are suggested by [BGHK92, GK92, Hal90].

According the objective view developed in [Kyb61], [Kyb74], and [Kyb97], constraints on the functions representing uncertainty are imposed by the statistical knowledge of the agent. These constraings may take either of two forms.

First, we may stipulate that the credibilities of the agent ought to form the lower envelope of a set of propositional probability functions which themselves are constrained by a priori or assumed statistical knowledge.

Second, we may stipulate that these credibilities be the corresponding intervals, where these intervals are based on inferred statistical knowledge.

Most writers who accept at all that we can have statistical knowledge will agree, I think, that the bounds imposed by this knowledge should be honored when they are relevant. (Even this, of course, requires a resolution of the problem of relevance: the reference class problem discussed in [Kyb83].)

What remains controversial, even among writers who agree so far, is the question of whether all statements in our object language can be tied to statistical background knowledge, or whether there are other, non-statistical sources of uncertainty. If there are other sources of uncertain knowledge, analogy, say, or similarity, it may still be the case that the inferential *import* of such sources of knowledge can only or best be expressed by statistical constraints on the metalinguistic level. Thus we might support an argument by analogy by the claim that among the worlds in which the analogical premises are true, a high proportion are also such that the analogical conclusion is true.

This is the thesis of combinatorial semantics, [Kyb97] which purports to be a general framework for the analysis of uncertain inference. Combinatorial semantics says it doesn't matter what the source of the uncertainty is — any uncertainty, at some level, can be looked on as statistical in character.

But statistical constraints, if they derive from empirical observation, can only be rough and ready, which is to say that what observation can warrant is the acceptance of a *set* of statistical statements as embodying our statistical knowledge about the world. Ordinarily this set will be convex in at least one parameter, such as $\{B_{n,p} : p \in [p_l, p_u]\}$ — the set of binomial distributions with parameter p in the interval $[p_l, p_u]$ — or $\{N(\mu, \sigma) : \mu \in [\mu_l, \mu_u] \wedge \sigma \in [\sigma_l, sigma_u]\}$ — a set of normal distributions similarly bounded. If our degrees of belief are to reflect our statistical knowledge, it is natural to suppose that they will be determined either directly by this approximate statistical knowledge, or by corresponding sets of probabilities. Even Bayesians will agree that the interval-valued constraints inherited from approximate statistical knowledge should be honored. (I leave to one side those extreme Bayesians who deny that we ever have even approximate statistical knowledge.)

And thus there is a degree of rapprochement between the generalized Bayesian view of uncertainties as given by convex sets of probability distributions, and the objectivistic view of uncertainties as determined by sets of statistical distributions known to hold in the world. Given some (approximate) knowledge about frequencies in the world, agents having that knowledge would be expected to constrain their rational credences by those frequencies. A natural way of representing these rational credences is by lower probabilities. Thus there is a close connection among a collection of views, ranging from that of Isaac Levi [Lev80], according to which rational credences of the agent are given by a convex set of probabilities over the sentences of the agent's language, to belief functions or the transferable belief model, in which it is the envelope itself, and not the fact that it is an envelope of probabilities that is important, to the view that takes the source of interval valued probabilities to be the (not necessarily convex) set of objective frequency distributions in the world that the agent can rationally claim to know.

It could be claimed that the various views we have mentioned reflect different ways of groping toward a realistic, approximate, objective, and evidence driven approach to uncertainty. If so, it is clear that we have not achieved that approach yet. But it remains, for many of us, an exciting goal toward which to strive.

References

[BGHK92] F. Bacchus, A.J. Grove, J.Y. Halpern, and D. Koller. From statistics to degrees of belief. In *AAAI-92*, pages 602–608, 1992.

[Car50] Rudolf Carnap. *The Logical Foundations of Probability*, volume Chicago. University of Chicago Press, 1950.

[Car80] Rudolf Carnap. A basic system for inductive logic: Part ii. In Jeffrey, editor, *Studies in Inductive Logic and Probability*, pages 7–155. University of California Press, Berkeley, 1980.

[Dem68] Arthur: Dempster. A generalization of bayesian inference. *Journal of the Royal Statistial Society Series B*, 30:205–247., 1968.

[Fin73] Terrence Fine. *Theories of Probability.* Academic Press, New York, 1973.

[Fin88] Terrence Fine. Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20:389–411, 1988.

[GK92] Joseph Y. Grove, Adam J.and Halpern and Daphne Koller. Random worlds and maximum entropy. In *7'th IEEE Symposium on Logic in Computer Science*, pages 22–33, 1992.

[Goo62] I. J. Good. Subjective probability as a measure of a non- measurable set. In Nagel, Suppes, and Tarski, editors, *Logic, Methodology and Philosophy of Science*, pages 319–329. University of California Press, Berkeley, 1962.

[Hal90] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

[Hem61] Carl G. Hempel. Deductive-nomological vs statistical explanation. In Herbert Feigl, editor, *Minnesota Studies in the Philosophy of Science III*, pages 98–169. University of Minnesota Press, 1961.

[Jay58] E. T. Jaynes. Probability theory in science and engineering. In *Colloquium Lectures in Pure and Applied Science*, volume 4, pages 152–187. Socony Mobile Oil Corp, Dallas Texas, 1958.

[Key52] John Maynard Keynes. *A Treatise on Probability.* Macmillan and Co., London, 1952.

[Koo40a] Bernard O. Koopman. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41:269–292, 1940.

[Koo40b] Bernard O. Koopman. The bases of probability. *Bulletin of the American Mathematical Society*, 46:763–774, 1940.

[Koo41] Bernard O Koopman. Intuitive probabilities and sequences. *Annals of Mathematics*, 42:169–187, 1941.

[Kyb61] Henry E. Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, 1961.

[Kyb74] Henry E. Kyburg, Jr. *The Logical Foundations of Statistical Inference Reidel.* Reidel, Dordrecht, 1974.

[Kyb83] Henry E. Kyburg, Jr. The reference class. *Philosophy of Science*, 50:374–397., 1983.

[Kyb87] Henry E. Kyburg, Jr. Bayesian and non-bayesian evidential updating. *A I Journal*, 31:271–294, 1987.

[Kyb88] Henry E. Kyburg, Jr. Probabilistic inference and probabilistic reasoning. In Shachter and Levitt, editors, *The Fourth Workshop on Uncertainty in Artificial Intelligence*, pages 221–228, 1988.

[Kyb97] Henry E. Kyburg, Jr. Combinatorial semantics: the semantics of frequent validity. *Computational Intelligence*, forthcoming, 1997.

[Lev80] Isaac Levi. *The Enterprise of Knowledge,*. MIT Press, Cambridge, 1980.

[Lew76] David K. Lewis. Probabilities of conditionals and conditional probabilities,. *Philosophical Review*, 85:297–315, 1976.

[Pop57] K.R. Popper. *The Propensity Interpretation of the Calculus of Probability and the Quantum Theory*, pages 65–70. 1957.

[Ram22] Frank P. Ramsey. Mr. keynes on probability. *The Cambridge Magazine*, 11, 1922.

[Ram31] Frank P. Ramsey. *Truth and Probability*, pages 156–198. Humanities Press, London, 1931.

[Sav54] L. J. Savage. *Foundations of Statistics*. John Wiley, New York, 1954.

[Sha76] Glenn Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, 1976.

[SK94] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[Sme88] Philippe Smets. The combination of evidence in the transferable belief model. Technical report, GinstitutionH, 1988.

[Smi61] C.A.B. Smith. Consistency in statistical inference and decision. *the Journal of the Royal Statistical Society B*, 23:1–37, 1961.

[Smi65] C.A.B. Smith. Personal probability and statistical analysis. *the Journal of the Royal Statistical Society A, 128*, pages 469–499, 1965.

[Wal91] Peter Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

[Wal95] Peter Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, B*, 57, 1995.

[Wal96] Peter Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, page forthcoming, 1996.