Indefinite Probabilities

for General Intelligence[1]

Matthew IKLÉ[a], Ben GOERTZEL[b], Izabela GOERTZEL[b]
[a]*Adams State College, Alamosa, Colorado*
and Novamente LLC.

[b]*Novamente LLC*

Abstract. The creation of robust mechanisms for uncertain inference is central to the development of Artificial General Intelligence systems. While probability theory provides a principled foundation for uncertain inference, the mathematics of probability theory has not yet been developed to the point where it is possible to handle every aspect of the uncertain inference process in practical situations using rigorous probabilistic calculations. Due to the need to operate within realistic computational resources, probability theory presently requires augmentation with heuristics in order to be pragmatic for general intelligence (as well as for other purposes such as large-scale data analysis).

The authors have been involved with the creation of a novel, general framework for pragmatic probabilistic inference in an AGI context, called Probabilistic Logic Networks (PLN). PLN integrates probability theory with a variety of heuristic inference mechanisms; it encompasses a rich set of first-order and higher-order inference rules, and it is highly flexible and adaptive, and easily configurable. This paper describes a single, critical aspect of the PLN framework, which has to with the quantification of uncertainty. In short, it addresses the question: *What should an uncertain truth value be, so that a general intelligence may use it for pragmatic reasoning?*

We propose a new approach to quantifying uncertainty via a hybridization of Walley's theory of imprecise probabilities and Bayesian credible intervals. This "indefinite probability" approach provides a general method for calculating the "weight-of-evidence" underlying the conclusions of uncertain inferences. Moreover, both Walley's imprecise beta-binomial model and standard Bayesian inference can be viewed mathematically as special cases of the more general indefinite probability model. Via exemplifying the use of indefinite probabilities in a variety of PLN inference rules (including exact and heuristic ones), we argue that this mode of quantifying uncertainty may be adequate to serve as an ingredient of powerful artificial general intelligence.

Introduction

As part of our ongoing work on the Novamente artificial general intelligence (AGI) system, we have developed a logical inference system called Probabilistic Logic Networks (PLN), designed to handle the various forms of uncertain inference that may confront a general intelligence -- including reasoning based on uncertain knowledge, and/or reasoning leading to uncertain conclusions (whether from certain or uncertain knowledge). Among the general high-level requirements underlying the development of PLN have been the following:

- To enable uncertainty-savvy versions of all known varieties of logical reasoning, including for

[1] The authors would like to thank Pei Wang for his very detailed and insightful comments on an earlier draft, which resulted in significant improvements to the paper.

instance higher-order reasoning involving quantifiers, higher-order functions, and so forth.

- To reduce to crisp "theorem prover" style behavior in the limiting case where uncertainty tends to zero.
- To encompass inductive and abductive as well as deductive reasoning.
- To agree with probability theory in those reasoning cases where probability theory, in its current state of development, provides solutions within reasonable calculational effort based on assumptions that are plausible in the context of real-world data.
- To gracefully incorporate heuristics not explicitly based on probability theory, in cases where probability theory, at its current state of development, does not provide adequate pragmatic solutions.
- To provide "scalable" reasoning, in the sense of being able to carry out inferences involving at least billions of premises. Of course, when the number of premises is fewer, more intensive and accurate reasoning may be carried out.
- To easily accept input from, and send input to, natural language processing software systems.

PLN implements a wide array of first-order and higher-order inference rules including (but not limited to) deduction, Bayes' Rule, unification, intensional and extensional inference, belief revision, induction, and abduction. Each rule comes with uncertain truth-value formulas, calculating the truth-value of the conclusion from the truth-values of the premises. Inference is controlled by highly flexible forward and backward chaining processes able to take feedback from external processes and thus behave adaptively.

The development of PLN has taken place under the assumption that probability theory is the "right" way to model uncertainty (more on this later). However, the mathematics of probability theory (and its interconnection with other aspects of mathematics) has not yet been developed to the point where it is feasible to use fully rigorous, explicitly probabilistic methods to handle every aspect of the uncertain inference process.

One of the major issues with probability theory as standardly utilized involves the very quantification of the uncertainty associated with statements that serve as premises or conclusions of inference. Using a single number to quantify the uncertainty of a statement is often not sufficient, a point made very eloquently by Wang ([1]), who argues in detail that the standard Bayesian approach does not offer any generally viable way to assess or reason about the "second-order uncertainty" involved in a given probability assignment. Probability theory provides richer mechanisms than this: one may assign a probability distribution to a statement, instead of a single probability value. But what if one doesn't have the data to fill in a probability distribution in detail? What is the (probabilistically) best approach to take in the case where a single number is not enough but the available data doesn't provide detailed distributional information? Current probability theory does not address this issue adequately. Yet this is a critical question if one wants to apply probability theory in a general intelligence context. In short, one needs methods of quantifying uncertainty at an intermediate level of detail between single probability numbers and fully known probability distributions. This is what we mean by the question: *What should an uncertain truth-value be, so that a general intelligence may use it for pragmatic reasoning?*

1. From Imprecise Probabilities to Indefinite Probabilities

2.

Walley's ([2]) theory of imprecise probabilities seeks to address this issue, via defining interval probabilities, with interpretations in terms of families of probability distributions. The idea of interval probabilities was originally introduced by Keynes ([3]), but Walley's version is more rigorous, grounded in the theory of envelopes of probability distributions. Walley's intervals, so-called "imprecise probabilities," are satisfyingly natural and consistent in the way they handle uncertain and incomplete information. However, in spite of a fair amount of attention over the years, this line of research has not yet been developed to the point of yielding robustly applicable mathematics.

Using a parametrized envelope of (beta-distribution) priors rather than assuming a single prior as would be typical in the Bayesian approach, Walley ([2], [4]) concludes that it is plausible to represent probabilities as intervals of the form $\left[\frac{m}{n+k}, \frac{m+k}{n+k}\right]$. In this formula, *n* represents the total number of observations, *m* represents the number of positive observations, and *k* is a parameter that Walley calls *s* and derives as a parameter of the beta distribution. Walley calls this parameter the learning parameter, while we will refer to it as the skepticism parameter. Note that the width of the interval of probabilities is inversely related to the number of observations *n*, so that the more evidence one has, the narrower the interval. The parameter *k* determines how rapidly this narrowing occurs. An interval of this sort is what Walley calls an "imprecise probability."

Walley's approach comes along with a host of elegant mathematics including a Generalized Bayes' Theorem. However it is not the only approach to interval probabilities. For instance, one alternative is Weichselberger's ([5]) axiomatic approach, which works with sets of probabilities of the form $[L,U]$ and implies that Walley's generalization of Bayes' rule is not the correct one.

One practical issue with using interval probabilities like Walley's or Weichselberger's in the context of probabilistic inference rules (such as those used in PLN) is the pessimism implicit in interval arithmetic. If one takes traditional probabilistic calculations and simplistically replaces the probabilities with intervals, then one finds that the intervals rapidly expand to [0,1]. This fact simply reflects the fact that the intervals represent "worst case" bounds. This same problem also affects Walley's and Weichselberger's more sophisticated approaches, and other approaches in the imprecise probabilities literature. The indefinite probabilities approach presented here circumvents these practical problems via utilizing interval probabilities that have a different sort of semantics – closely related to, but not the same as, those of Walley's interval probabilities.

Indefinite probabilities, as we consider them here, are represented by quadruples of the form ([L,U],b,k) – thus, they contain two additional numbers beyond the [L,U] interval truth values proposed by Keynes, and one number beyond the ([L,U],k) formalism proposed by Walley. The semantics involved in assigning such a truth value to a statement S is, roughly, "I assign a probability of b to the hypothesis that, after I have observed k more pieces of evidence, the truth value I assign to S will lie in the interval [L,U]." In the practical examples presented here we will hold k constant and thus will deal with truth value triples ([L,U],b).

The inclusion of the value b, which defines the credibility level according to which [L,U] is a credible interval (for hypothesized future assignments of the probability of S, after observing k more pieces of evidence), is what allows our intervals to generally remain narrower than those produced by existing imprecise probability approaches. If b=1, then our approach essentially reduces to imprecise probabilities, and in pragmatic inference contexts tends to produce intervals [L,U] that approach [0,1]. The use of b<1 allows the inferential production of narrower intervals, which are more useful in a real-world inference context.

In practice, to execute inferences using indefinite probabilities, we make heuristic distributional assumptions, assuming a "second order" distribution which has [L,U] as a (100*b)% credible interval, and then "first order" distributions whose means are drawn from the second-order distribution. These distributions are to be viewed as heuristic approximations intended to estimate unknown probability values existing in hypothetical future situations. The utility of the indefinite probability approach may be dependent on the appropriateness of the particular distributional assumptions to the given application situation. But in practice we have found that a handful of distributional forms seem to suffice to cover common-sense inferences (beta and bimodal forms seem good enough for nearly all cases; and here we will only give examples covering the beta distribution case).

Because the semantics of indefinite probabilities is different from that of ordinary probabilities, or imprecise probabilities, or for example NARS truth values, it is not possible to say objectively that any one of these approaches is "better" than the other one, as a mathematical formalism. Each approach is better than the others at mathematically embodying its own conceptual assumptions. From an AGI perspective, the value of an approach to quantifying uncertainty lies in its usefulness when integrated with a pragmatic probabilistic reasoning engine. The bulk of this paper will be concerned with showing how indefinite probabilities behave when incorporated in the logical reasoning rules utilized in the PLN inference framework, a component of the Novamente AI Engine. While complicated and dependent on many factors, this is nevertheless the sort of evaluation that we consider most meaningful.

Section 2 deals with the conceptual foundations of indefinite probabilities, clarifying their semantics in the context of Bayesian and frequentist philosophies of probability. Section 3 outlines the pragmatic computational method we use for doing probabilistic and heuristic inference using indefinite probabilities. Section 4 presents a number of specific examples involving using indefinite probabilities within single inference steps within the PLN inference framework.

3. The Semantics of Uncertainty

4.

The main goal of this paper is to present indefinite probabilities as a pragmatic tool for uncertain inference, oriented toward utilization in AGI systems. Before getting practical, however, we will pause in this section to discuss the conceptual, semantic foundations of the "indefinite probability" notion. In the course of developing the indefinite probabilities approach, we found that the thorniest aspects lay not in the mathematics or software implementation, but rather in the conceptual interpretation of the truth values and their roles in inference.

In the philosophy of probability, there are two main approaches to interpreting the meaning of probability values, commonly labeled frequentist and Bayesian ([6]). There are many shades of meaning to each interpretation, but the essential difference is easy to understand. The frequentist approach holds that a probability should be interpreted as the limit of the relative frequency of an event-category, calculated over a series of events as the length of the series tends to infinity. The subjectivist or Bayesian approach holds that a probability should be interpreted as the degree of belief in a statement, held by some observer; or in other words, as an estimate of how strongly an observer believes the evidence available to him supports the statement in question. Early proponents of the subjectivist view were Ramsey ([7]) and de Finetti ([8]), who argued that for an individual to display self-consistent betting behavior they would need to assess degrees of belief according to the laws of probability theory. More recently Cox's Theorem ([9]) and related mathematics ([10]) have come into prominence as providing a rigorous foundation for subjectivist probability. Roughly speaking, this mathematical work shows that if the observer assessing subjective probabilities is to be logically consistent, then their plausibility estimates must obey the standard rules of probability.

From a philosophy-of-AI point of view, neither the frequentist nor the subjectivist interpretations, as commonly presented, is fully satisfactory. However, for reasons to be briefly explained here, we find the subjectivist interpretation more acceptable, and will consider indefinite probabilities within a subjectivist context, utilizing relative frequency calculations for pragmatic purposes but giving them an explicitly subjectivist rather than frequentist interpretation.

The frequentist interpretation is conceptually problematic in that it assigns probabilities only in terms of limits of sequences, not in terms of finite amounts of data. Furthermore, it has well-known difficulties with the assignment of probabilities to unique events that are not readily thought of as elements of ensembles. For instance, what was the probability, in 1999, of the statement S holding that "A great depression will be brought about by the Y2K problem"? Yes, this probability can be cast in terms of relative frequencies in various ways. For instance, one can define it as a relative frequency across a set of hypothetical "possible worlds": across all possible worlds similar to our own, in how many of them did the Y2K problem bring about a great depression? But it's not particularly natural to assume that this is what an intelligence must do in order to assign a probability to S. It would be absurd to claim that, in order to assign a probability to S, an intelligence must explicitly reason in terms of an ensemble of possible worlds. Rather, the claim must be that whatever reasoning a mind does to evaluate the probability of S may be implicitly interpreted in terms of possible worlds. This is not completely senseless, but is a bit of an irritating conceptual stretch.

The subjectivist approach, on the other hand, is normally conceptually founded either on rational betting behaviors or on Cox's Theorem and its generalizations, both of which are somewhat idealistic.

No intelligent agent operating within a plausible amount of resources can embody fully self-consistent betting behavior in complex situations. The irrationality of human betting behavior is well known; to an extent this is due to emotional reasons, but there are also practical limitations on the complexity of the situation in which any finite mind can figure out the correct betting strategy.

And similarly, it is too much to expect any severely resource-constrained intelligence to be fully self-consistent in the sense that the assumptions of Cox's theorem require. In order to use Cox's Theorem to justify the use of probability theory by practical intelligences, it seems to us, one would need to take another step beyond Cox, and argue that if an AI system is going to have a "mostly sensible" measure of plausibility (i.e. if its deviation from Cox's axioms are not too great), then its intrinsic plausibility measure must be *similar to* probability. We consider this to be a viable line of argument, but will pursue this point in another paper – to enlarge on such matters here would take us too far afield.

Walley's approach to representing uncertainty is based explicitly on a Bayesian, subjectivist interpretation; though whether his mathematics has an alternate frequentist interpretation is something he has not explored, to our knowledge. Similarly, our approach here is to take a subjectivist perspective on the foundational semantics of indefinite probabilities (although we don't consider this critical to our approach; quite likely it could be given a frequentist interpretation as well.) Within our basic subjectivist interpretation, however, we will frequently utilize relative frequency calculations when convenient for pragmatic reasoning. This is conceptually consistent because within the subjectivist perspective, there is still a role for relative frequency calculations, so long as they are properly interpreted.

Specifically, when handling a conditional probability $P(A|B)$, it may be the case that there is a decomposition $B=B_1+...+B_n$ so that the B_i are mutually exclusive and equiprobable, and each of $P(A|B_i)$ is either 0 or 1. In this case the laws of probability tell us $P(A|B) = P(A|B_1) P(B_1| B) + ... + P(A|B_n) P(B_n|B) = (P(A|B_1) + ... + P(A|B_n))/n$, which is exactly a relative frequency. So, in the case of statements that are decomposable in this sense, the Bayesian interpretation implies a relative frequency based interpretation (but not a "frequentist" interpretation in the classical sense). For decomposable statements, plausibility values may be regarded as the means of probability distributions, where the distributions may be derived via subsampling (sampling subsets C of $\{B_1,...,B_n\}$, calculating $P(A|C)$ for each subset, and taking the distribution of these values; as in the statistical technique known as bootstrapping). In the case of the "Y2K" statement and other similar statements regarding unique instances, one option is to think about decomposability across possible worlds, which is conceptually controversial.

4.1.Indefinite Probability

4.2.

We concur with the subjectivist maxim that a probability can usefully be interpreted as an estimate of the plausibility of a statement, made by some observer. However, we suggest introducing into this notion a more careful consideration of the role of evidence in the assessment of plausibility. We introduce a distinction that we feel is critical, between

- the ordinary (or "definite") plausibility of a statement, interpreted as the degree to which the evidence already (directly or indirectly) collected by a particular observer supports the statement.
- the "indefinite plausibility" of a statement, interpreted as the degree to which the observer believes that the overall body of evidence potentially available to him supports the statement.
-

The indefinite plausibility is related to the ordinary plausibility, but also takes into account the potentially limited nature of the store of evidence collected by the observer at a given point in time. While the ordinary plausibility is effectively represented as a single number, the indefinite plausibility is more usefully represented in a more complex form. We suggest to represent an indefinite plausibility as a quadruple $([L,U],b,k)$, which when attached to a statement S has the semantics "I assign an ordinary plausibility of b to the statement that 'Once k more items of evidence are collected, the ordinary plausibility of the statement S will lie in the interval [L,U]'". Note that indefinite plausibility is thus defined as "second order plausibility" – a plausibility of a plausibility.

As we shall see in later sections of the paper, for most computational purposes it seems acceptable to leave the parameter k in the background, assuming it is the same for both the premises and the conclusion of an inference. So in the following we will mainly speak of indefinite probabilities as $([L,U],b)$ triples, for sake of simplicity. The possibility does exist, however, that in future work, inference algorithms will be designed that utilize k explicitly.

Now, suppose we buy the Bayesian argument that ordinary plausibility is best represented in terms of probability. Then it follows that indefinite plausibility is best represented in terms of second-order probability, i.e. as "I assign probability b to the statement that 'Once k more items of evidence have been collected, the probability of the truth of S based on this evidence will lie in the interval $[L,U]$'".

4.2.1.An Interpretation in Term of Betting Behavior

4.2.2.

To justify the above definition of indefinite probability more formally, one approach is to revert to betting arguments of the type made by de Finetti in his work on the foundations of probability. As will be expounded below, for computational purposes, we have taken a pragmatic frequentist approach, based on underlying distributional assumptions. However, for purposes of conceptual clarity, a more subjectivist de Finetti style justification is nevertheless of interest. So, in this subsection, we will describe a "betting scenario" that leads naturally to a definition of indefinite probabilities.

Suppose we have a category C of discrete events, e.g. a set of tosses of a certain coin which has heads on one side and tails on the other.

Next, suppose we have a predicate S, which is either True or False (boolean values) for each event within the above event-category. C. For example, if C is a set of tosses of a certain coin, then S could be the event "Heads". S is a function from events into Boolean values.

If we have an agent A, and the agent A has observed the evaluation of S on n different events, then we will say that n is the amount of evidence that A has observed regarding S; or we will say that A has made n observations regarding S.

Now consider a situation with three agents: the House, the Gambler, and the Meta-gambler.

As the name indicates, the House is going to run a gambling operation, involving generating repeated events in category C, and proposing bets regarding the outcome of future events in C.

More interestingly, House is also going to propose bets to the Meta-gambler, regarding the behavior of the Gambler.

Specifically, suppose the House behaves as follows.

After the Gambler makes n observations regarding S, House offers Gambler the opportunity to make what we'll call a "de Finetti" type bet regarding the outcome of the next observation of S.That is, House offers Gambler the opportunity:

"You must set the price of a promise to pay $1 if the next observation of S comes out True, and $0 if there it does not. You must commit that I will be able to choose either to buy such a promise from you at the price you have set, or require you to buy such a promise from me. In other words: you set the odds, but I decide which side of the bet will be yours."

Assuming the Gambler does not want to lose money, the price Gambler sets in such a bet, is the "operational subjective probability" that Gambler assigns that the next observation of S will come out True.

As an aside, House might also offer Gambler the opportunity to bet on sequences of observations, e.g. it might offer similar "de Finetti" price-setting opportunities regarding predicates like "The next 5 observations of S made will be in the ordered pattern (True, True, True, False, True)." In this case, things become interesting if we suppose Gambler thinks that: For each sequence Z of {True, False} values emerging from repeated observation of S, any permutation of Z has the same (operational subjective) probability as Z. Then, Gambler thinks that the series of observations of S is "exchangeable", which means intuitively that S's subjective probability estimates are really estimates of the "underlying probability of S being true on a random occasion." Various mathematical conclusions follow from the assumption that Gambler does not want to lose money, combined with the assumption that Gambler believes in exchangeability.

Next, let's bring Meta-gambler into the picture. Suppose that House, Gambler and Meta-gambler have all together been watching n observations of S. Now, House is going to offer Meta-gambler a special opportunity. Namely, he is going to bring Meta-gambler into the back room for a period of time. During this period of time, House and Gambler will be partaking in a gambling process involving the predicate S.

Specifically, while Meta-gambler is in the back room, House is going to show Gambler k new observations of S. Then, after the k'th observation, House is going to come drag Meta-gambler out of the back room, away from the pleasures of the flesh and back to the place where gambling on S occurs.

House then offers Gambler the opportunity to set the price of yet another de-Finetti style bet on yet another observation of S.

Before Gambler gets to set his price, though, Meta-gambler is going to be given the opportunity of placing a bet regarding what price Gambler is going to set.

Specifically, House is going to allow Meta-gambler to set the price of a de Finetti style bet on a proposition of Meta-gambler's choice, of the form:

Q = "Gambler is going to bet an amount p that lies in the interval [L,U]"

For instance Meta-gambler might propose

"Let Q be the proposition that Gambler is going to bet an amount lying in [.4, .6] on this next observation of S. I'll set at 30 cents the price of a promise defined as follows: To pay $1 if Q comes out True, and $0 if it does not. I will commit that you will be able to choose either to buy such a promise from me at this price, or require me to buy such a promise from you."

I.e., Meta-Gambler sets the price corresponding to Q, but House gets to determine which side of the bet to take.

Let us denote the price set by Meta-gambler as b; and let us assume that Meta-gambler does not want to lose money.

Then, b is Meta-gambler's subjective probability assigned to the statement that:

"Gambler's subjective probability for the next observation of S being True lies in [L,U]."

But, recall from earlier that the indefinite probability

<L,U,b,k>
attached to S means that:

"The estimated odds are b that after k more observations of S, the estimated probability of S will lie in [L,U]"

or in other words

"[L,U] is a b-level credible interval for the estimated probability of S after k more observations."

In the context of an AI system reasoning using indefinite probabilities, there is no explicit separation between the Gambler and the Meta-gambler; the same AI system makes both levels of estimate. But this is of course not problematic, so long as the two components
(first-order probability estimation and b-estimation) are carried out separately.

One might argue that this formalization in terms of betting behavior doesn't really add anything practical to the indefinite probabilities framework as already formulated. At minimum, however, it does make the relationship between indefinite probabilities and the classical subjective interpretation of probabilities quite clear.

4.2.3.A Pragmatic Frequentist Interpretation

4.2.4.

Next, it is not hard to see how the above-presented interpretation of an indefinite plausibility can be provided with an alternate justification in relative frequency terms, in the case where one has a statement S that is decomposable in the sense described above. Suppose that, based on a certain finite amount of evidence about the frequency of a statement S, one wants to guess what one's frequency estimate will be once one has seen a lot more evidence. This guessing process will result in a probability distribution across frequency estimates – which may itself be interpreted as a frequency via a "possible worlds" interpretation. One may think about "the frequency, averaged across all possible worlds, that we live in a world in which the observed frequency of S after k more observations will lie in interval I." So, then, one may interpret ([L,U],b, N) as meaning "b is the frequency of possible worlds in which the observed frequency of S, after I've gathered k more pieces of evidence, will lie in the interval [L,U]."

This interpretation is not as conceptually compelling as the betting-based interpretation given above – because bets are real things, whereas these fictitious possible worlds are a bit slipperier. However, we make use of this frequency-based interpretation of indefinite probabilities in the practical computational implementation of indefinite probability presented in the following sections -- without, of course, sacrificing the general Bayesian interpretation of the indefinite probability approach. In the end, we consider the various interpretations of probability to be in the main complementary rather than contradictory, providing different perspectives on the same very useful mathematics.

Moving on, then: To adopt a pragmatic frequency-based interpretation of the second-order plausibility in the definition of indefinite plausibility, we interpret "I assign probability b to the statement that 'Once k more items of evidence are collected, the probability of the truth of S based on this evidence will lie in the interval [L,U]'" to mean "b is the frequency, across all possible worlds in which I have gathered k more items of evidence about S, of worlds in which the statement 'the estimated probability of S lies in the interval [L,U]' is true ". This frequency-based interpretation allows us to talk about a probability distribution consisting of probabilities assigned to values of 'the estimated probability of S', evaluated across various possible worlds. This probability distribution is what, in the later sections of the paper, we call the "second-order distribution." For calculational purposes, we assume a particular distributional form for this second-order distribution.

Next, for the purpose of computational implementation, we make the heuristic assumption that the statement S under consideration is decomposable, so that in each possible world, "the estimated probability of S" may be interpreted as the mean of a probability distribution. For calculational purposes, in our current implementation we assume a particular distributional form for these probability distributions, which we refer to as "the first-order distributions."

The adoption of a frequency-based interpretation for the second-order plausibility seems hard to avoid if one wants to do practical calculations using the indefinite probabilities approach. On the other hand, the adoption of a frequency-based interpretation for the first-order plausibilities is an avoidable convenience, which is appropriate only in some situations. We will discuss below how the process of reasoning using indefinite probabilities can be simplified, at the cost of decreased robustness, in cases where decomposability of the first order probabilities is not a plausible assumption.

So, to summarize, in order to make the indefinite probabilities approach computationally tractable, we begin by restricting attention to some particular family D of probability distributions. Then, we interpret an interval probability attached to a statement as an assertion that: "There is probability *b* that the subjective probability of the statement, after I have made k more observations, will appear to be drawn from a distribution with a mean in this interval."

Then, finally, given this semantics and a logical inference rule, one can ask questions such as: "If each of the premises of my inference corresponds to some interval, so that there is probability *b* that after k more observations the distribution governing the premise will appear to have a mean in that interval; then, what is an interval so that *b* of the family of distributions of the conclusion have means lying in that interval?"[2] We may then give this final interval the interpretation that, after k more observations, there is a probability b that the conclusion of the inference will appear to lie in this final interval. (Note that, as mentioned above, the parameter k essentially "cancels out" during inference, so that one doesn't need to explicitly account for it during most inference operations, so long as one is willing to assume it is the same in the premises and the conclusion.)

.

In essence, this strategy merges the idea of imprecise probabilities with the Bayesian concept of credible intervals; thus the name "indefinite probabilities" ("definite" having the meaning of "precise," but also the meaning of "contained within specific boundaries" – Walley's probabilities are contained within specific boundaries, whereas ours are not).

5. Varieties of Uncertain Truth Value

6.

We have discussed above the semantic foundations of truth values of the form $([L,U],b,k)$. Before proceeding further to discuss inference with these objects, it is worth pausing to note that, within the PLN inference framework, this is one of several forms of truth value object utilized. In general, the forms of truth value objects currently utilized in PLN are:

- IndefiniteTruthValues (as already discussed)
- SimpleTruthValues (to be described below)
- DistributionalTruthValues (which, in their most general form, involve maintaining a whole probability distribution of possible probability distributions attached to a statement)

SimpleTruthValues take two, equivalent forms:

[2] In fact, this is a minor oversimplification, because the credibility value *b* may be different for each premise and for the conclusion.

-
-
- (s,w), where s is the mean strength, the best single-number probability assignable to the statement; w is the "weight of evidence", a number in [0,1] that tells you, qualitatively, how much you should believe the strength estimate
- (s,n), where s is the mean strength; n is the "count", a number > 0 telling you, qualitatively, the total amount of evidence that was evaluated in order to assess s

Borrowing from Pei Wang's NARS system ([11], [12]), the weight-of-evidence and count in SimpleTruthValues are related via $w=\frac{n}{n+k1}$, where $k1$ is the system-wide skepticism parameter. Weight-of evidence is essentially a normalized version of the count. Similarly, n and [L,U] may be interrelated by a heuristic formula, n = k1 (1-W)/W where W=U-L is the width of the credible interval. More rigorous formulas interrelating n and ([L,U],b,k) may be formulated if one makes specific distributional assumptions (e.g. assuming an underlying Bernoulli process), but these lead to qualitatively similar results to the above heuristic formula, and will be discussed in a later publication.

Note also that if [L,U] is of the form [m/(n+k1), (m+k1)/(n+1k)], then k1 is interpretable as the number of hypothetical additional pieces of evidence being considered to obtain the mean-estimates in [L,U]. That is, in this case the parameter k1 associated with the ([L,U],b) truth value may be interpreted as equal to the parameter k used in the definition of indefinite truth values.

As hinted above, however, the above descriptions mask the complexity of the actual truth-value objects (except in the SimpleTruthValue case, which actually is simple). In the indefinite probabilities approach, in practice, each IndefiniteTruthValue object is also endowed with three additional parameters:

- an indicator of whether [L,U] should be considered as a symmetric or asymmetric credible interval.
- a family of "second-order" distributions, used to govern the second-order plausibilities described above.
-
- a family of "first-order" distributions, used to govern the first-order plausibilities described above
  -

Combined with these additional parameters, each truth-value object essentially provides a compact representation of a single second-order probability distribution with a particular, complex structure.

7. Inference Using Indefinite Probabilities

8.

8.1.

We now describe the general process according to which a PLN inference formula is evaluated using indefinite probabilities. This process contains three basic steps. We will first present these steps in an abstract way; and then, in the following section, exemplify them in the context of specific probabilistic inference formulas corresponding to inference rules like term logic and modus ponens deduction, Bayes' Rule, and so forth.

In general, the process is the same for any inference formula that can be thought of as taking in premises labeled with probabilities, and outputting a conclusion labeled with a probability. The examples given here will pertain to single inference steps, but the same process may be applied holistically to a series of inference steps, or most generally an "inference tree" of inference steps, each building on conclusions of prior steps.

Step One in the indefinite probabilistic inference process is as follows. Given intervals, $\left[L_i, U_i\right]$, of mean premise probabilities, we first find a distribution from the "second-order distribution family" supported on $\left[L1_i, U1_i\right]$É$\left[L_i, U_i\right]$, so that these means have $\left[L_i, U_i\right]$ as $(100 \cdot b_i)$% credible intervals. The intervals $\left[L1_i, U1_i\right]$ are either of the form $\left[\frac{m}{n+k}, \frac{m+k}{n+k}\right]$, when the "interval-type" parameter associated with the premise is asymmetric; or are such that both of the intervals $\left[L1_i, L_i\right]$ and $\left[U_i, U1_i\right]$ each have probability mass $b_i/2$, when interval-type is symmetric.

Next, in Step Two, we use Monte-Carlo methods on the set of premise truth-values to generate final distribution(s) of probabilities as follows. For each premise, we randomly select n_1 values from the ("second-order") distribution found in step 1. These n_1 values provide the values for the means for our first-order distributions. For each of our n_1 first-order distributions, we select n_2 values to represent the first-order distribution. We apply the applicable inference rule (or tree of inference rules) to each of the n_2 values for each first-order distribution to generate the final distribution of first-order distributions of probabilities. We calculate the mean of each distribution and then – in Step Three of the overall process -- find a $(100 \cdot b_i)$% credible interval, $\left[L_f, U_f\right]$, for this distribution of means.

When desired, we can easily translate our final interval of probabilities, $\left[L_f, U_f\right]$, into a final, $\left(s_f, n_f, b_f\right)$ triple of strength, count, and credibility levels, as outlined above.

Getting back to the "Bayesianism versus frequentism" issues raised in the previous section, if one wished to avoid the heuristic assumption of decomposability regarding the first-order plausibilities involved in the premises, one would replace the first-order distributions assumed in Step Two with Dirac delta functions, meaning that no variation around the mean of each premise plausibility would be incorporated. This would also yield a generally acceptable approach, but would result in overall narrower conclusion probability intervals, and we believe would in most cases represent a step away from realism and robustness.

8.1.1.The Procedure in More Detail

We now run through the above three steps in more mathematical detail.

In Step One of the above procedure, we assume that the mean strength values follow some given initial probability distribution $g_i(s)$ with support on the interval $[L1_i, U1_i]$. If the interval-type is specified as asymmetric, we perform a search until we find values *k2* so that $\int_{L_i}^{U_i} g_i(s)\,ds=b$, where $L_i=\frac{m_i}{n_i+k2}$ and $U_i=\frac{m_i+k2}{n_i+k2}$. If the interval-type is symmetric, we first ensure, via parameters, that each first-order distribution is symmetric about its mean s_i, set $L_i=s_i-d$, $U_i=s_i+d$ and perform a search for d to ensure that $\int_{L_i}^{U_i} g_i(s)\,ds=b$. In either case, each of the intervals $[L_i, U_i]$ will be a $(100\cdot b)\%$ credible interval for the distribution $g_i(s)$.

We note that we may be able to obtain the appropriate credible intervals for the distributions $g_i(s)$ only for certain values of b. For this reason, we say that a value b is truth-value-consistent whenever it is feasible to find $(100\cdot b)\%$ credible intervals of the appropriate type.

In Step Two of the above procedure, we create a family of distributions, drawn from a pre-specified set of distributional forms, and with means in the intervals $[L_i, U_i]$. We next apply Monte Carlo search to form a set of randomly chosen "premise probability tuples". Each tuple is formed via selecting, for each premise of the inference rule, a series of points drawn at random from randomly chosen distributions in the family. For each randomly chosen premise probability tuple, the inference rule is executed. And then, in Step Three, to get a probability value s_f for the conclusion, we take the mean of this distribution. Also, we take a credible interval from this final distribution, using a pre-specified credibility level b_f, to obtain an interval for the conclusion $[L_f, U_f]$.

When the interval-type parameter is set to asymmetric, to find the final count value n_f, note that $m=ns$. Then, since the skepticism parameter *k1* is a system parameter, we need only solve the following equation for the count n_f:

$$\int_{\frac{n_f s_f}{n_f+k1}}^{\frac{n_f s_f+k1}{n_f+k1}} g_f(s)\,ds=b$$

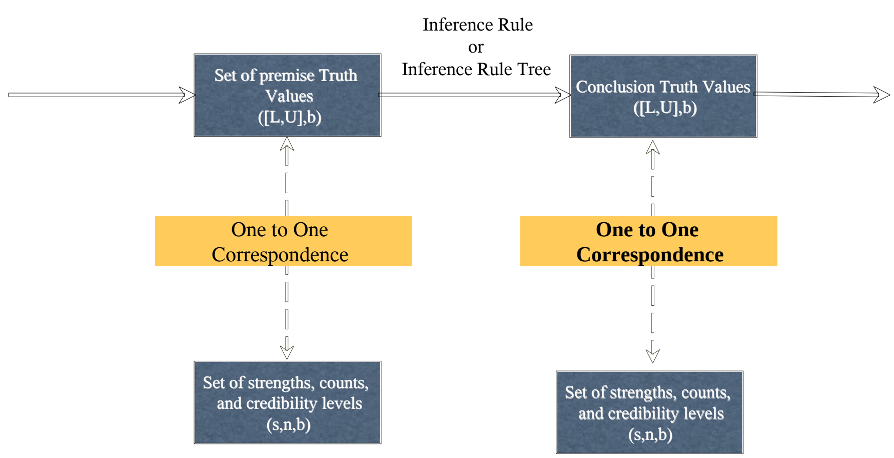

Diagram 1

For symmetric interval-types, we use the heuristic $n = c(1\text{-}W)/W$ where W, the interval width, is $W=U_f - L_f$.

9. A Few Detailed Examples

10.

In this section we report some example results obtained from applying the indefinite probabilities approach in the context of simple inference rules, using both symmetric and asymmetric interval-types.

Comparisons of results on various inference rules indicated considerably superior results in all cases when using the symmetric intervals. As a result, we will report results for five inference rules using the symmetric rules; while we will report the results using the asymmetric approach for only one example (term logic deduction).

10.1.

10.2.A Detailed Bet-Binominal Example for Bayes' Rule

First we will treat Bayes' Rule, a paradigm example of an uncertain inference rule -- which however is somewhat unrepresentative of inference rules utilized in PLN, due to its non-heuristic, exactly probabilistic nature.

The beta-binomial model is commonly used in Bayesian inference, partially due to the conjugacy of the beta and binomial distributions. In the context of Bayes' rule, Walley develops an imprecise beta-binomial model (IBB) as a special case of an imprecise Dirichlet model (IDM). We illustrate our indefinite probabilities approach as applied to Bayes' rule, under the same assumptions as these other approaches.

We treat here the standard form for Bayes' rule: $P(A|B)=\dfrac{P(A)P(B|A)}{P(B)}$.

We consider the following simple example problem. Suppose we have 100 gerbils of unknown color; 10 gerbils of known color, 5 of which are blue; and 100 rats of known color, 10 of which are blue. We wish to estimate the probability of a randomly chosen blue rodent being a gerbil.

The first step, in our approach, is to obtain initial probability intervals. We obtain the following sets of initial probabilities shown in Tables 1-3, corresponding to credibility levels b of 0.95, and 0.982593, respectively.

Table 1. Intervals for Crediblity Level 0.90

EVENT	[L, U]	[L1, U1]

G	$\left[\frac{10}{21},\frac{12}{21}\right]$ =[.476190, 0.571429]	[0.419724, 0.627895]
R	$\left[\frac{8}{21},\frac{12}{21}\right]$ =[0.380952, 0.571429]	[0.26802, 0.684361]
B|G	[0.3, 0.7]	[0.0628418, 0.937158]
B|R	[0.06, 0.14]	[0.0125683, 0.187432]

Table 2. Intervals for Crediblity Level 0.95

EVENT	**[L, U]**	**[L1, U1]**
G	$\left[\frac{10}{21},\frac{12}{21}\right]$ =[.476190, 0.571429]	[0.434369, 0.61325]
R	$\left[\frac{8}{21},\frac{12}{21}\right]$ =[0.380952, 0.571429]	[0.29731, 0.655071]
B|G	[0.3, 0.7]	[0.124352, 0.875649]
B|R	[0.06, 0.14]	[0.0248703, 0.17513]

We begin our Monte Carlo step by generating n_1 random strength values, chosen from Beta distributions proportional to $x^{ks-1}(1-x)^{k(1-s)-1}$ with mean values of $s=\frac{11}{21}$ for *P*(G); $s=\frac{10}{21}$ for *P*(R); $s=\frac{1}{2}$ for $P(B|G)$; and $s=\frac{1}{10}$ for $P(B|R)$, and with support on [L1, U1]. Each of these strength values then serve, in turn, as parameters of standard Beta distributions. We generate a random sample of n_2 points from each of these standard Beta distributions.

We next apply Bayes' Theorem to each of of the $n_1 n_2$ quadruples of points, generating n_1 sets of sampled distributions. Averaging across each distribution then gives a distribution of final mean strength values. Finally, we transform our final distribution of mean strength values back to (*s, n, b*) triples.

10.2.1.Experimental Results

10.2.2.

Results of our Bayes' Rule experiments are summarized in Tables 3-6.

Table 3. Final Probability Intervals for P(G|B) using initial b-values of 0.90	
CREDIBILITY LEVEL	**INTERVAL**
0.90	[0.715577, 0.911182]
0.95	[0.686269, 0.924651]

Table 4. Final Average Strength and Count Values using initial b-values of 0.90		
CREDIBILITY LEVEL	**STRENGTH**	**COUNT via n=k(1-w)/w**
0.90	0.832429	41.1234
0.95	0.832429	31.9495

Table 5. Final Probability Intervals for P(G|B) using initial b-values of 0.95	
CREDIBILITY LEVEL	**INTERVAL**
0.90	[0.754499, 0.896276]
0.95	[0.744368, 0.907436]

Table 6. Final Average Strength and Count Values using initial b-values of 0.95		
CREDIBILITY LEVEL	**STRENGTH**	**COUNT**
0.90	0.835557	60.5334
0.95	0.835557	51.3239

10.2.3.

10.2.4. Comparison to Standard Approaches

It is not hard to see, using the above simple test example as a guide, that our indefinite probabilities approach generalizes both classical Bayesian inference and Walley's IBB model. First note that single distributions can be modeled as envelopes of distributions with parameters chosen from uniform distributions. If we model *P*(B) as a uniform distribution; *P*(A) as a single beta distribution and $P(B|A)$ as a single binomial distribution, then our method reduces to usual Bayesian inference. If, on the other hand, we model *P*(B) as a uniform distribution; *P*(A) as an envelope of beta distributions; and $P(B|A)$ as an envelope of binomial distributions, then our envelope-approach reduces to Walley's IBB model. Our envelope-based thus approach allows us to model P(B) by any given family of distributions, rather than restricting us to a uniform distribution. This allows for more flexibility in accounting for known, as well as unknown, quantities.

To get a quantitative comparison of our approach with these others, we modeled the above test example using standard Bayesian inference as well as Walley's IBB model. To carry out standard Bayesian analysis, we note that given that there are 100 gerbils whose blueness has not been observed, we are dealing with 2^{100} "possible worlds" (i.e. possible assignments of blue/non-blue to each gerbil). Each of these possible worlds has 110 gerbils in it, at least 5 of which are blue, and at least 5 of which are non-blue.

For each possible world w, we can calculate the probability that drawing 10 gerbils from the population of 110 existing in world W yields an observation of 5 blue gerbils and 5 non-blue gerbils. This probability may be written $P(D|H)$, where D is the observed data (5 blue and 5 non-blue gerbils) and H is the hypothesis (the possible world W).

Applying Bayes' rule, we have $P(H|D)=\frac{P(D|H)P(H)}{P(D)}$. Assuming that $P(H)$ is constant across all possible worlds, we find that $P(H|D)$ is proportional to $P(D|H)$. Given this distribution for the possible values of the number of blue gerbils, one then obtains a distribution of possible values $P(gerbil|blue)$, and calculates a credible interval. The results of this Bayesian approach are summarized in Table 7.

Table 7. Final Probability Intervals for P(G|B)

CREDIBILITY LEVEL	INTERVAL
0.90	[0.8245614035, 0.8630136986]
0.95	[0.8214285714, 0.8648648649]

We also applied Walley's IBB model to our example, obtaining (with k=10) the interval $\left[\frac{2323}{2886},\frac{2453}{2886}\right]$, or approximately [0.43007, 0.88465]. In comparison, the hybrid method succeeds at maintaining narrower intervals, albeit at some loss of credibility.

With a k-value of 1, on the other hand, Walley's approach yields an interval of [0.727811, 0.804734]. This interval may seem surprising since it does not include the average given by Bayes' theorem. However, it is sensible given the logic of Walley's approach. In this approach, we assume no prior knowledge of P(G) and we have 10 new data points in support of the proposition that $P(G|B)$ is 55/65. So we assume beta distribution priors with s=0 and s=1 for the "endpoints" of P(G) and use n=10 and p=11/13 for the binomial distribution for $P(G|B)$.

The density function thus has the form:

$$f(x)=\frac{x^{ks}(1-x)^{k(1-s)}x^{np}(1-x)^{n(1-p)}}{\int_0^1 x^{ks}(1-x)^{k(1-s)}x^{np}(1-x)^{n(1-p)}\,dx}$$

Now for s=1 and value of the learning parameter k=1, the system with no prior knowledge starts with the interval [1/3,2/3]. With only 10 data points in support of p=11/13 and prior assumption of no knowledge or prior p=1/2, Walley's method is (correctly according to its own logic) reluctant to move quickly in support of p=11/13, without making larger intervals via larger k-values.

10.3.

10.4.A Detailed Beta-Binominal Example for Deduction

Next we consider another inference rule, term logic deduction, which is more interesting than Bayes' Rule in that it combines probability theory with an heuristic independence assumption. The independence-assumption-based PLN deduction rule, as derived in ([13]), has the following form, for "consistent" sets *A*, *B*, and *C*[3]:

$$s_{AC} = s_{AB} s_{BC} + \frac{(1-s_{AB})(s_C - s_B s_{BC})}{1-s_B}$$

where

$$s_{AC} = P(C|A) = \frac{P(AÇC)}{P(A)}$$

assuming the given data $s_{AB} = P(B|A)$, $s_{BC} = P(C|B)$, $s_A = P(A)$, $s_B = P(B)$, and $s_C = P(C)$.

Our example for the deduction rule will consist of the following premise truth-values. In the table, we provide the values for [*L*, *U*], and *b*, as well as the values corresponding to the mean *s* and count *n*.

Table 8. Premise Truth-Values Used for Deduction with Symmetric Intervals

Premise	***s***	**[*L*,*U*]**	**[*L1, U1*]**
A	11/23	[10/23, 12/23]≈[0.434783, 0.521739]	[0.383226, 0.573295]
B	0.45	[0.44, 0.46]	[0.428142, 0.471858]
AB	0.413043	[0.313043, 0.513043]	[0.194464, 0.631623]
BC	8/15	[7/15, 9/15] ≈ [0.466666, 0.6]	[0.387614, 0.679053]

We now vary the premise truth-value for variable C, keeping the mean s_C constant, in order to study changes in the conclusion count as the premise width [L,U] varies. In the table below, *b* = 0.9 for both premise and conclusion, and s_C = 0.59. The final count n_{AC} is found via the heuristic formula $n_{AC} = k(1-W)/W$.

Table 9. Deduction Rule Results Using Symmetic Intervals

Premise *C*		**Conclusion *AC***		
[*L*,*U*]	**[*L1, U1*]**	**sAC**	**[*L*,*U*]**	***nAC***

[3] Consistency here is defined in terms of a set of inequalities interrelating the premise probabilities, which are equivalent to the requirement that the conclusion according to the formula given here must lie in the interval [0,1].

[0.44, 0.74]	[0.262131, 0.917869]	0.575514	[0.434879, 0.68669]	24.0004
[0.49, 0.69]	[0.371421, 0.808579]	0.571527	[0.478815, 0.650717]	48.1727
[0.54, 0.64]	[0.48071, 0.69929]	0.571989	[0.52381, 0.612715]	102.478
[0.58, 0.60]	[0.568142, 0.611858]	0.571885	[0.4125, 0.6756]	197.892

For comparison of using symmetric intervals versus asymmetric, we also tried identical premises for the means using the asymmetric interval approach. In so doing, the premise intervals [*L*, *U*] and [*L1*, *U1*] are different as shown in the table below, using b = 0.9 as before

Table 10. Premise Truth-Values Used for Deduction with Asymmetric Intervals			
Premise	***s***	**[*L*,*U*]**	**[*L1, U1*]**
A	11/23	[0.44, 0.52]	[0.403229, 0.560114]
B	0.45	[0.44, 0.462222]	[0.42818, 0.476669]
AB	0.413043	[0.38, 0.46]	[0.3412, 0.515136]
BC	8/15	[0.48, 0.58]	[0.416848, 0.635258]

Table 11. Deduction Rule Results Using Symmetic Intervals				
Premise *C*		**Conclusion *AC***		
[*L*,*U*]	**[*L1, U1*]**	**sAC**	**[*L*,*U*]**	***nAC***
[0.40, 0.722034]	[0.177061, 0.876957]	0.576418	[0.448325, 0.670547]	35
[0.45, 0.687288]	[0.28573, 0.801442]	0.577455	[0.461964, 0.661964]	40
[0.50, 0.652543]	[0.394397, 0.725928]	0.572711	[0.498333, 0.628203]	67
[0.55, 0.617796]	[0.526655, 0.600729]	0.568787	[0.4125, 0.6756]	125

10.4.1.Modus Ponens

Another important inference rule is Modus Ponens, which is the form of deduction standard in predicate logic rather than term logic. Term logic deduction as described above is preferable from an uncertain inference perspective, because it generally supports more certain conclusions. However, the indefinite probabilities approach can also handle Modus Ponens, it simply tends to assign conclusions fairly wide interval truth-values.

The general form of Modus Ponens is:

A

A → B

| -

B

To derive an inference formula for this rule, we reason as follows. Given that we know P(A) and P(B | A), we know nothing about P(B | ¬A). Hence P(B) lies in the interval [Q,R] = [P(A and B), 1- (P(A) - P(A and B))] = [P(B|A) P(A), 1 - P(A) + P(B|A)P(A)].

For the Modus Ponens experiment reported here we used the following premises: For A, we used $\left([L,U],b\right)=\left(\left[\frac{10}{23},\frac{12}{23}\right],0.9\right)$; and for A → B, we used ([L,U],b)=([0.313043,0.513043],0.9). We proceed as usual, choosing distributions of distributions for both P(A) and P(B | A). Combining these we find a distribution of distributions [Q,R] as defined above. Once again, by calculating means, we end up with a distribution of [Q,R] intervals. Finally, we find an interval [L,U] that contains ($100 \cdot b$)% of the final [Q,R] intervals. In our example, our final [L,U] interval at the b = 0.9 level is [0.181154, 0.736029].

10.5.Conjunction

10.6.

Next, the AND rule in PLN uses a very simple heuristic probabilistic logic formula:

$$P(A\ \mathrm{AND}\ B)=P(A)\,P(B)$$

To exemplify this, we describe an experiment consisting of assuming a truth-value of ([L,U],b)=([0.4,0.5],0.9) for A and a truth-value of ([L,U],b)=([0.2,0.3],0.9).

The conclusion truth-value for $P(A$ AND $B)$ then becomes. ([L,U],b)=([0,794882,0.123946],0.9)

10.7.Revision

The final inference rule we study in this paper, the "revision" rule, is used to combine different estimates of the truth-value of the same statement. Very sophisticated approaches to belief revision are possible within the indefinite probabilities approach; for instance we are currently exploring the possibility of integrating entropy optimization heuristics as described in ([14]) into PLN for this purpose ([15]). At present, however, we are using a relatively simple heuristic approach for revising truth-values. This approach seems to be effective in practice, although it lacks the theoretical motivation of the entropy minimization approach.

Suppose D1 is the second-order distribution for premise 1 and D2 is the second-order distribution for D2. Suppose further that n_1 is the count for premise 1 and n_2 is the count for premise 2. Let $w_1 = n_1/(n_1+n_2)$ and $w_2 = n_2/(n_1+n_2)$ and then form the conclusion distribution D = w_1 D1 + w_2D2. We then generate our [L,U],b truth-value as usual.

As an example, consider the revision of the following two truth-values ([0.1, 0.2], 0.9) and ([0.3, 0.7], 0.9). Estimating the counts using $n = k1(1\text{-}W)/W$ gives count values of 90 and 15 respectively. Fusing the two truth-values yields ([0.19136, 0.222344], 0.9) with a resulting count value of 79.1315.

11. Conclusions

Using credible intervals of probability distribution envelopes to model second-order plausibiltiies, we have generalized both Walley's imprecise probabilities model and the standard Bayesian model into a novel "indefinite probabilities" approach. On Bayes' rule, the results obtained via this method fall squarely between standard Bayesian models and Walley's interval model, providing more general results than Bayesian inference, while avoiding the quick degeneration to worst-case bounds inherent with imprecise probabilities. On more heuristic PLN inference rules, the indefinite probability approach gives plausible results for all cases attempted, as exemplified by the handful of examples presented in detail here.

Comparing the merits of the indefinite probability approach with those of other approaches such as Walley's imprecise probabilities, the standard Bayesian approach, or NARS is difficult, for conceptual rather than practical or mathematical reasons. Each of these approaches proceeds from subtly different theoretical assumptions. To a great extent, each one is the correct approach according to its own assumptions; and the question then becomes which assumptions are most useful in the context of pragmatic uncertain inferences. In other words: the ultimate justification of a method of quantifying uncertainty, from an AGI or narrow-AI perspective, lies in the quality of the inferential conclusions that can be obtained using it in practice.

However, the drawing of such conclusions is not a matter of uncertainty quantification alone – it requires an inference engine (such as PLN), embedded in a systematic framework for feeding appropriate data and problems to the inference engine (such as the Novamente AI Engine). And so, the assessment of indefinite probabilities from this perspective goes far beyond the scope of this paper. What we have done here is to present some initial, suggestive evidence that the indefinite probabilities approach may be useful for artificial general intelligence systems – firstly because it rests on a sound conceptual and semantic foundation; and secondly because when applied in the context of a variety of PLN inference rules (representing modes of inference hypothesized to be central to AGI), it consistently gives intuitively plausible results, rather than giving results that intuitively seem too indefinite (like the intervals obtained from Walley's approach, which too rapidly approach [0,1] after inference), or giving results that fail to account fully for premise uncertainty (which is the main issue with the standard, Bayesian or frequentist, first-order-probability approach).

References

[1] Pei Wang, "The Limitation of Bayesianism, Artificial Intelligence," 158(1), 97-106, 2004.

[2] Peter Walley, "Statistical reasoning with imprecise probabilities," London ; New York : Chapman and Hall, 1991.

[3] John Maynard Keynes, A Treatise on Probability, Dover Books, 1921.2004

[4]

Kurt Weichselberger, Thomas Augustin: "On the Symbiosis of Two Concepts of Conditional Interval Probability," ISIPTA 2003: 606-

[5] Peter Walley, "Inferences from Multinomial Data: Learning about a Bag of Marbles," Journal of the Royal Statistical Society. Series B (Methodological), Vol. 58, No. 1, pp. 3-57, (1996).

[6] Stanford Encyclopedia of Philosophy, "Interpretations of Probability," 2003, http://plato.stanford.edu/entries/probability-interpret/

[7] Frank Ramsey, Foundations of Mathematics, 1931

[8] Bruno de Finetti, Theory of Probability, (translation by AFM Smith of 1970 book) 2 volumes, New York: Wiley, 1974-5.

[9] R. T. Cox, "Probability, Frequency, and Reasonable Expectation," Am. Jour. Phys., 14, 1-13, (1946).

[10] Michael Hardy, Advances in Applied Mathematics, August 2002, pages 243-292

[11] Pei Wang, Rigid Flexibility: The Logic of Intelligence, Springer, 2006.

[12] Pei Wang, Non-Axiomatic Reasoning System: Exploring the essence of intelligence, Ph.D. Dissertation, Indiana University, 1995.

[13] Ben Goertzel, Matthew Iklé, Izabela Freire Goertzel, Ari Heljakka, Probabilistic Logic Networks: A General Framework for Applied Uncertain Inference, in preparation, expected publication 2007.

[14] Gabriele Kern-Isberner and Wilhelm Rödder. Belief revision and information fusion on optimum entropy, International Journal of Intelligent Systems, Volume 19, Issue 9 , 2004, Pages 837 – 857

[15] Ben Goertzel, Matthew Iklé, "Revision of Indefinite Probabilities via Entropy Minimization", in preparation, expected publication 2007.