(2)

AD-A255 471

REPORT DO

Form Approved
OPM No. 0704-0188

ublic reporting burden for this collection of information
maintaining the data needed, and reviewing the collecti
for reducing this burden, to Washington Headquarters
the Office of Information and Regulatory Affairs, Office

ng instructions, searching existing data sources gathering and
ir aspect of this collection of information, including suggestions
Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	Unknown	Unknown

4. TITLE AND SUBTITLE

Probabilistic Inference

5. FUNDING NUMBERS

DAAB10-86-C-0567

6. AUTHOR(S)

Kyburg, Henry E., Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Rochester
Department of Philosophy
Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army CECOM Signals Warfare Directorate
Vint Hill Farms Station
Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

92-TRF-0050

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Statement A: Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words) It is important to distinguish probabilistic reasoning from probabilistic inference. Probabilistic reasoning may concern the manipulation of knowledge of probabilities in the context of decision theory, or it may involve the updating of probabilities in the light of new evidence via Bayes' theorem or some other procedure. Both of these operations are essentially dedecutive in character. Contrasted with these procedures of manipulating or computing with probabilities, is the use of probabilistic rules of inference: rules that lead from one sentence (or a set of sentences) to another sentence, but do so in a way that need not be truth preserving. One could attempt to get along without probabilistic inference in AI, but it would be very difficult and unnatural. Instances of such rules are several classes of inference rules associated with statistics, and some rules discussed by philosophers. In artificial intelligence the rules that fall into this category are (mainly) default rules; these are not generally construed probabilstically; but obviously default rules that more often lead you astray than to the truth would be poor ones.

14. SUBJECT TERMS

Artificial Intelligence, Data Fusion, Probabilistic Inference, Probabilistic Reasoning

15. NUMBER OF PAGES

10

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

1987

Probabilistic Inference

by Henry E. Kyburg, Jr.

University of Rochester

1. probabilistic inference and probabilistic reasoning.

Uncertainty enters into human reasoning and inference in at least two distinct ways. One way concerns choices among alternative actions. For good reasons, having to do with Dutch books (Ramsey 1950), this kind of uncertainty is associated with the classical probability axioms. It is this form of uncertainty that is used in computing the expectations that are fed into decision rules. It has been argued that the most general and useful form of representation for these uncertainties is that of a convex set of classical probability functions, defined over an algebra of propositions (Levi 1980, Kyburg 1987). Such a representation includes as special cases belief functions and most interval representations of uncertainty. Manipulating these probability representations, together with utility functions, constitutes one form of probabilistic reasoning.

In addition to merely representing uncertainty and employing it in decision theory, we are concerned with how uncertainties are modified or updated in response to evidence. The classical way of doing this, for classical probabilities, is by means of Bayes' theorem: if statement E becomes known, is accepted as evidence, then the new or updated probability P' of any statement H in our algebra becomes the likelihood of E multiplied by the ratio of the old probability of H to the old probability of E:

$$P'(H) = P(H/E) * (P(H)/P(E))$$

This is called 'conditionalization'. Conditionalization can be extended to the more general approach that represents uncertainty by convex sets of classical probabilities: it can be shown that if each classical probability function in a convex set of probability functions is updated by conditionalizing on the evidence E, the result will be a new convex set of classical probability functions, provided E does not have zero probability on all the original probability functions (Kyburg, 1987).

There are other ways in which one might want to update probabilities than by conditionalization -- certain forms of direct inference, in which probabilities are derived from knowledge of statistics or chances, have been shown to conflict with conditionalization, for example (Levi, 1980). But while any of these procedures have a perfect right to be called 'probabilistic reasoning,' they are not what I mean by probabilistic inference.

In inference in general, one begins with certain statements or propositions (representations of states of affairs), premises, and goes through a process that leads to another statement, the conclusion. In ordinary deductive logic, the process is such as to preserve truth: if the premises are true, so is the conclusion. Note that the probabilistic reasoning mentioned above fits this deductive pattern. From "tosses of this coin are independent and heads occurs half the time," we infer, not probabilistically, but deductively, that triples of tosses consisting of three heads occur an eighth of the time.

What is controversial is whether or not there is any form of inference other than deductive inference. Is there any way of arguing from premises to conclusion that is not truth preserving, and if there is, why would one want to do it anyway? Of course there is a tradition that considers "inductive inference," "ampliative inference," and the like

(Kneale and Kneale, 1962). But this is a tradition in philosophy that many regard as a bit musty, and so we will approach the question from the other side, from the direction of artificial intelligence.

There, the answer is clear: this form of inference is what non-monotonic logics (for example) are designed to capture. Since the inferences do not preserve truth, we have to be able to back up: if we enlarge the premises, we may have to shrink the conclusions. Non-monotonic inference is not generally taken to be probabilistic, but work on non-monotonic logic suggests that there is interest in inference rules -- that is, rules that lead from premises to the acceptance of a conclusion -- that need not be truth-preserving. Many people want to be able to detach conclusions from their premises. (Not all approaches to non-monotonic logic allow full detachment; de Kleer's ATMS (de Kleer, 1986), for example, requires that tags reflecting the assumptions used in carrying out an inference be carried along with the conclusions.)

2. why accept?

Despite the fact that some people are interested in non-deductive inference, we may still sensibly ask why they should be: why should we accept any statements that are not (say) mathematical or logical truths? It might be thought that we couldn't use conditionalization for updating without acceptance: after all, when we up-date on evidence E, we take the probability of E to be 1. And once a statement has a probability of 1 (or of 0) that probability can never be changed by conditionalization. But there are other ways to handle up-dating: Jeffrey's rule (Jeffrey, 1965), for example, or various net-propagation procedures, such as Pearl's (Pearl, 1986).

In principle, there is no reason that human or machine knowledge in a

certain domain should not be represented by a complete algebra of statements and a probability distribution (or a set of probability distributions) over them, in which no empirical statement ever receives a probability of 0 or 1. Such a system would have no need for a probabilistic rule of inference.

As a matter of practicality, except for the most trivial domains, the idea does not seem feasible at all. Our empirical scientific knowledge is expressed, not in probabilities (for the most part) but in categorical statements. There is a sense in which we may want to say that our science is uncertain; but there is no probability we associate with the principle that the vector sum of the forces acting on a static body must be zero. We do not take measurement to result in statements such as "with probability .9, the reading 4.30 was obtained," nor do we report the result of the measurement as an unbounded normal probability distribution.

No one, I suspect, has ever tried to represent a significant piece of knowledge or expertise in this way. It would be perverse. When we measure a rod by a method M whose distribution of error is normal with a mean of zero and a standard deviation of .01, we don't worry about the finite probability that the reading is off by more than .05. As for the distribution of error itself, we don't even keep the data: the hypothesis was confirmed well enough. Maybe the mean is really 10^{-6} rather than 0. Maybe the variance isn't exactly .01. But the probability of a significant deviation is too small to bother about. This is probabilistic inference in action.

In testing a statistical hypothesis, the standard goal is to devise a rule that will erroneously reject that hypothesis no more than α of the time. Such a test will lead you to a false rejection no more frequently

than α (Lehman, 1959). Of course α is a free parameter; but we choose α to be small enough that the possibility of making this sort of error does not worry us. The size we choose reflects how seriously we take the mistake in question. If it is very serious, we want to be very sure (but we can't ask for a guarantee) that it won't happen. It is very bad form in a particular case in which a hypothesis has been tested and rejected to say that the probability is at most alpha that it was falsely rejected. (But as Birnbaum has pointed out (1969), while we can learn not to say this, it is hard to know what else to think.) For present purposes we leave aside whatever other desiderata we might want to take account of in designing tests for statistical hypotheses.

Or consider the simplest and most elegant of all forms of statistical inference: you have a normally distributed quantity $\underline{X}$, but you don't know the parameters of its distribution. Nevertheless, since it is normally distributed, you know the distribution of the quantity $\underline{t} = (\underline{N})^{1/2}(\bar{\underline{x}} - \mu$ $((\underline{s}^{-1})$, where $\underline{x}$ and $\underline{s}$ are the sample mean and standard deviation, and is the unknown population mean. Knowing the distribution of $\underline{t}$, you can therefore compute the probability, for example, that

$$\bar{\underline{x}} - \underline{ts}/\underline{N}^{1/2} < \mu < \bar{\underline{x}} - \underline{ts}/\underline{N}^{1/2}$$

If you pick some probability level that makes you comfortable under the circumstances, and you are indifferent between over and under-estimating μ, then you will have an exact interval estimate of the unknown mean μ, indexed by $\underline{fp}$, a level of fiducial probability or practical certainty.

Or consider the most common form of confidence interval inference: you have a binomial population with an unknown parameter $\underline{r}$; you draw a sample from the population, and observe a relative frequency $\underline{f}$; you construct a class of intervals $(\underline{p}_l, \underline{p}_u)$ such that whatever the true value

of r may be, the probability is at least p that the sample frequency will fall in the corresponding interval. We infer, after observing the sample, that the sample fell in its representative interval. But it will have done this if and only if r lies between a certain maximum and a certain minimum value. These values determine what is called a confidence interval, and in particular, a 100p% confidence interval, since its limits require the specification of an acceptable p.

Outside of statistics, consider Levi (1967). Levi is concerned with the circumstances under which one ought to add a hypothesis to one's corpus of knowledge. The famous Rule A for doing so involves, in addition to the probability of the hypothesis, and a measure of the epistemic content of the hypothesis, and a further parameter q, which varies from 0 to 1 and functions as an index of caution.

In artificial intelligence Matthew Ginsberg (1985) applies a technique much like that of binomial confidence interval inference (the main difference being that he uses a rougher approximation) to the problem of inferring an interval characterizing the reliability of a default rule in non-monotonic logic. In order to do this, he finds it necessary to introduce a parameter g, which he calls "gullibility".

Finally, in my own work (1961, 1974) I have adopted a "purely probabilistic" rule of acceptance. That is, a body of knowledge is indexed by a "level of acceptance"; statements whose probabilities (relative to a body of knowledge of even higher rank) are greater than this level of acceptance may be accepted.

3. probabilistic acceptance

The simplest idea is just to accept those statements whose probability exceeds a certain critical number. This number may have to be

changed to reflect different circumstances -- it will be context dependent -- but so, we may suppose, are α, g, q, p, and fp context dependent.

In what way is acceptance level context dependent? One natural answer is that acceptance level depends on what is, or might be expected to be, at stake. If the range of stakes that we are contemplating is limited -- for example, it can't be more than 10 to 1 -- then probabilities greater than .9 are indistinguishable (behaviorally) from probabilities of 1, and probabilities less than .1 are indistinguishable from probabilities of 0.

It also follows from these considerations that probabilities larger than the level of acceptance, or smaller than 1 - the level of acceptance, are just not significant as probabilities. That is, it makes no sense to bet at odds of 1000:1 on a statement that gets its probability from a statistical statement whose acceptance level is only .99. The constraint cuts both ways.

Most of the acceptance rules mentioned above run afoul of the lottery paradox (Kyburg, 1961). That is, each of a set of statements Si (e.g., "ticket i will not win the lottery") may be probable enough to be accepted, and at the same time may jointly contradict other accepted statements (e.g., "there will be a winner."). The only exception is the acceptance principle advocated by Levi, which links acceptance to expected epistemic utility; only statements demonstrably consistent with what you have already accepted are candidates for future acceptance.

How serious the lottery paradox is depends on what other machinery you have. It is not deadly if you limit yourself to a probabilistic rule of acceptance. It will follow that any logical consequence of a single statement in your corpus of knowledge should also be in it; but it will

not follow that every consequence of the set of sentences in your corpus of knowledge will also be in it. The latter would indeed lead to a hopeless sort of inconsistency. The former would not. If the size of the lottery is adjusted to my level of acceptance, I will answer your question about whether ticket i will win with a categorical "no." But I will answer your question of whether its true that neither ticket i nor ticket $i + 1$ will win by saying, "I don't know."

This seems not unreasonable. Or look at the matter in another way: given a (deductive) argument from premisses P_1 ... P_2 to a conclusion C, consider when the argument obligates you to accept C. It seems natural to say that more is required than merely that each of the premises be accepted; I must also be willing to accept the conjunction of the premises.

Even this feature might be advantageous in AI. There is surely an epistemic difference between a conclusion reached in one step from a single premise, and a conclusion that requires a number of premises. This difference disappears if the acceptability of the single premise of the first argument is no greater than that of the conjunction of all the premises in the second argument. A purely probabilistic rule of acceptance automatically reflects this fact.

4. conclusion

It is important to distinguish probabilistic reasoning from probabilistic inference. Probabilistic reasoning may concern the manipulation of knowledge of probabilities in the context of decision theory, or it may involve the updating of probabilities in the light of new evidence via Bayes' theorem or some other procedure. Both of these operations are essentially deductive in character.

Contrasted with these procedures of manipulating or computing with probabilities, is the use of probabilistic rules of inference: rules that lead from one sentence (or a set of sentences) to another sentence, but do so in a way that need not be truth preserving. One could attempt to get along without probabilistic inference in AI, but it would be very difficult and unnatural.

Instances of such rules are several classes of inference rules associated with statistics, and some rules discussed by philosophers. In artificial intelligence the rules that fall into this category are (mainly) default rules; these are not generally construed probabilstically, but obviously default rules that more often led you astray than to the truth would be poor ones.

The simplest probabilitic rules of inference -- a high probability rules -- has some curious consequences, but it does not seem that these consequences need interfere with the useful application of the rule.

BIBLIOGRAPHY

Birnbaum, Alan (1969) "Concepts of Statistical Evidence," in Morgenbesser, et al. (eds.), Philosophy Science and Method, St. Martin's Press, New York.

de Kleer, J. (1986) "An Assumption-Based ATMS," AI Journal 28, 1986, 127-162.

Ginsberg, Matthew (1985), "Does Prob...", IJCAI 85.

Kneale, W., and Kneale, M. (1962) The Development of Logic, Oxford.

Kyburg, H.E., Jr. (1961) Probability and the Logic of Ratiional Belief Wesleyan University Press, Middletown CT.

_____, (1974) The Logical Foundations of Statistical Inference.

_____, (1987) "Bayesian and non-Bayesian Evidential Updating," AI Journal, forthcoming.

Lehman, E.L. (1959) Testing Statistical Hypotheses, John Wiley & Sons, New York.

Levi, Isaac (1967) Gambling with Truth, Knopf, New York.

_____ (1980) The Enterprise of Knowledge, MIT Press, Cambridge, MA.

Pearl, Judea (1986) "Fusion, Propagation, and Structuring in Belief Networks," AI Journal 29, 1986, pp. 241-288.

Ramsey, F.P. (1950) The Foundations of Mathematics and Other Essays, Humanities Press, New York.