The Limitation of Bayesianism

Pei Wang

Department of Computer and Information Sciences
Temple University, Philadelphia, PA 19122
pei.wang@temple.edu

Abstract

In the current discussion about the capacity of Bayesianism in reasoning under uncertainty, there is a conceptual and notational confusion between the explicit condition and the implicit condition of a probability evaluation. Consequently, the limitation of Bayesianism is often seriously underestimated. To represent the uncertainty of a belief system where revision is needed, it is not enough to assign a probability value to each belief.

Key words: Bayes' Theorem, conditioning, Jeffrey's rule, revision, updating, ignorance, sensitivity.

1 Introduction

In recent years, Bayesian networks have achieved great success. It has been applied to various problems, and taken by more and more people as a normative theory of reasoning, both for the human mind, and for artificial intelligence systems.

Though the Bayesian approach is indeed a powerful tool for many theoretical and practical problems, in the current study its limitation is often seriously underestimated, due to a conceptual and notational confusion. The problem was first addressed in Wang (1993), but it has got little attention, and the confusion continues to spread. This research note provides a more focused and comprehensive discussion on this issue.

According to Pearl (1990), traditional Bayesianism is defined by the following attributes:

- willingness to accept subjective belief as an expedient substitute for raw data,

- reliance on complete (i.e., coherent) probabilistic models of beliefs,
- adherence to Bayes' conditionalization as the primary mechanism for updating belief in light of new information.

When probability theory is applied in a reasoning system, it usually starts by assuming a proposition space S, which contains all the propositions that the system can represent and process. S is normally generated from a set of atomic propositions, using logical operators NOT ($\neg$), AND ($\wedge$), and OR ($\vee$). A probability distribution P is defined on S, and for every proposition $x \in S$, its probability evaluation $P(x)$ is a real number in [0, 1].[1] The function P satisfies the following axioms [Kolmogorov (1950)]:

- $P(x \vee \neg x) = 1$.
- $P(x \vee y) = P(x) + P(y)$, if $y \in S$ and $x \wedge y$ is false.

For any x and y in S, the probability of x under the condition that y is true is a conditional probability evaluation $P(x|y) = P(x \wedge y)/P(y)$. From it we get Bayes' Theorem

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{1}$$

Though the above mathematical definitions and results are acknowledged by all people using probability theory for reasoning, the Bayesian approach interprets them differently. According to Bayesianism (as defined above), the probability of a proposition h in a system is the system's degree of belief on h, according to certain background knowledge K (or call it experience, data, evidence, and so on).

The system starts with a *prior* probability distribution P_0, determined by background knowledge K_0 at time t_0. At time t_1, when a piece of new knowledge e is collected, Bayes' Theorem is applied to change P_0 into a *posterior* probability distribution P_1, where the probability of a proposition h is

$$P_1(h) = P_0(h|e) = \frac{P_0(e|h)P_0(h)}{P_0(e)} \tag{2}$$

Now P_1 is based on K_1, which includes both K_0 and e. By repeatedly applying

[1] Sometimes a probability distribution is not defined on a proposition space S, but on a set of events or a set of possible values of certain random variables. However, since the probability of an event E is the same as the probability of the proposition "E happens", and the probability of a random variable X to have value v is the same as probability of the proposition "$X = v$", for the current discussion it is fine to use the proposition space representation.

Bayes' Theorem in this "conditioning" process, the system learns new knowledge, and adjusts its beliefs accordingly [Heckerman (1999); Pearl (2000)].

According to the above description, we see that under the Bayesian interpretation, a probabilistic evaluation $P(h)$ is always "conditional", in the sense that it is not an objective property of the proposition h, but a relation between h and background knowledge K. For this reason, the previous inference rule can be written as

$$P_{K_1}(h) = P_{K_0}(h|e) = \frac{P_{K_0}(e|h)P_{K_0}(h)}{P_{K_0}(e)} \tag{3}$$

where K_0 and K_1 are the background knowledge the system has at time t_0 and t_1, respectively. In the following, I will call them "implicit conditions" of the corresponding probability distribution function, because they are conditions of the probability functions, and they are usually implicitly assumed in the formula.

A common practice is to represent the dependency of a probability to an implicit condition as a conditional probability, that is, to write equation (3) as

$$P(h|K_1) = P(h|e \wedge K_0) = \frac{P(e|h \wedge K_0)P(h|K_0)}{P(e|K_0)} \tag{4}$$

Such a representation can be found in many publications, for example, Cheeseman (1985), Heckerman (1999), and Pearl (1988, 2000).

Since in a conditional probability the condition is explicitly represented, in the following, I will call them "explicit conditions". The argument this research note makes is that, in general, it is improper to represent an implicit condition as an explicit condition, and that their difference shows a serious limitation of Bayesianism, which is related to several previous debates on related topics.

2 Explicit conditions vs. implicit conditions

Since Bayesian learning is carried out by equation (2), the knowledge the system can directly learn in this way must be represented as an explicit condition e. This means:

(1) It is a (binary) proposition (otherwise it cannot be in S).
(2) It is in S (otherwise its probability $P_0(e)$ is undefined).
(3) $P_0(e) > 0$ (otherwise it cannot be used as a denominator).

The first restriction should not be confused with that a random variable X can have more than two possible values. Though X can have more than two values, the knowledge "$X = v_i$" is still a binary proposition.

These restrictions are not unknown — for example, a similar list is discussed in Diaconis and Zabell (1983) and Pearl (1990). Since all learning methods have their restrictions, it is not a surprise that Bayesian conditioning cannot learning everything. However, here the problem is that the above restrictions are not applied to the implicit conditions of probability distribution functions:

(1) An implicit condition may include statistical conclusions and subjective probabilistic estimates, which are not binary propositions.
(2) An implicit condition only needs to be related to S, but not necessarily in S. For example, "Tweety is a bird and cannot fly" can be part of an implicit condition, even though S includes only "Birds fly", and does not include the name "Tweety" at all.
(3) Even if a proposition is assigned a prior probability of zero according to one knowledge source, it is still possible for the proposition to be assigned a non-zero probability according to another knowledge source.

Now we can see that only certain types of implicit conditions can be represented as explicit conditions. It follows that if some knowledge is not available when the prior probability is determined, it is impossible to be put into the system through Bayesian conditioning. We cannot assume that we can always start with a "non-informative" prior probability distribution, and learn the relevant knowledge when it becomes available.

Therefore, it is wrong to represent an implicit condition as an explicit condition, and the previous equations (3) and(4) are not equivalent to each other. Though both equations are correct, they have different meanings.

From a practical point of view, the three restrictions are not trivial, since they mean that although the background knowledge can be probabilistic-valued, all new knowledge must be binary-valued; no novel concept and proposition can appear in new knowledge; and if a proposition is given a probability 1 or 0, such a belief cannot be changed in the future, no matter what happens. We could build such a system, but unfortunately it would be a far cry from the everyday reasoning process of a human being.

Some people claim that the Bayesian approach is sufficient for reasoning with uncertainty, and many people treat Bayes' Theorem as a generally applicable learning rule, because explicit conditions and implicit conditions of a probability evaluation are seldom clearly distinguished in related discussions. Without such a distinction, the illusion arises that all the knowledge supporting a probability distribution function can be represented by explicit conditions, and can therefore be learned by the system using Bayes' conditionalization.

3 Revision vs. updating

Within the Bayesian tradition, there is a way to handle new evidence that is not a binary proposition. After a prior probability distribution P_0 is assigned to a proposition space S, some new evidence may show that "The probability of proposition e ($e \in S$) should be changed to p" (i.e., $P_1(e) = p$). In this situation, assuming the conditional probabilities that with e or $\neg e$ as explicit condition are unchanged (i.e., $P_1(h|e) = P_0(h|e)$), we can update the probability evaluation for every proposition h in S to get a new distribution function according to Jeffrey's rule [Jeffrey (1965); Diaconis and Zabell (1983); Kyburg (1987); Pearl (1988)]:

$$P_1(h) = P_0(h|e) \times p + P_0(h|\neg e) \times (1 - p) \tag{5}$$

If we interpret "e happens" as "e's probability should be changed to 1", then Bayes' Theorem, when used for learning as in (2), becomes a special case of Jeffrey's rule, where $p = 1$. [2]

A related method was suggested to process uncertain evidence e, where a "virtual proposition" v is introduced to represent the new knowledge as "a (unspecified) proposition v is true, and $P_0(e|v) = p$" [Cheeseman (1986); Pearl (1988)]. Then a new conditional probability distribution can be calculated (after considering the new knowledge) in the following way:

$$P_1(h) = P_0(h|v) = P_0(h|e \wedge v) \times P_0(e|v) + P_0(h|\neg e \wedge v) \times P_0(\neg e|v) \tag{6}$$

Under the assumption that $P_0(x|e \wedge v) = P_0(x|e)$ and $P_0(x|\neg e \wedge v) = P_0(x|\neg e)$, equation (6) can be reduced into (7):

$$P_1(h) = P_0(h|v) = P_0(h|e) \times p + P_0(h|\neg e) \times (1 - p) \tag{7}$$

Therefore we end up with Jeffrey's rule. The only difference is that here the prior probability is not *updated* directly, but is instead *conditionalized* by a virtual condition (the unspecified proposition v). However, no matter which procedure is followed and how the process is interpreted, the result is the same [Pearl (1990)].

Some other systems process uncertain evidence by providing *likelihood ratios* of virtual propositions [Pearl (1988); Heckerman (1988)]. This method also leads to conditionalization of a virtual condition, therefore is semantically equivalent to the previous approach [Pearl (1990)].

---

[2] Jeffrey's rule can be seen as a special case of "Probability Kinematics" [Jeffrey (1965)], by which $P_1(h) = \sum_{i=1}^{n} P_0(h|e_i)P_1(e_i)$, where $\sum_{i=1}^{n} P_1(e_i) = 1$.

Jeffrey's rule (and its equivalent forms) avoids the first restriction of Bayes' Theorem, that is, the new evidence must be a binary proposition. Also, if conditional probability is directly defined by de Finetti's coherent conditional probability (that is, not as $P(x|y) = P(x \wedge y)/P(y)$), it is possible to do conditioning on an event which has prior probability 0 [Coletti et al. (1993)]. Furthermore, Pearl defines "Neo-Bayesianism" by adding structural information (i.e., the topological structure of a Bayesian network) into tradition Bayesianism. With this kind of information, conditional probability can be introduced independent of the absolute probability values, and, therefore, the above limitations of conditioning is overcome [Pearl (1990)].

Though the above methods are well justified, they only cover a special case. In general, by "revision" (or "learning", "belief change", and so on), I mean the process by which a system changes the degree of belief (no matter what they are called and defined) of certain proposition h from $B(h) = p_1$ to $B(h) = p_2$, according to evidence e. By "updating", in this paper I mean a special case of the above process where e takes the form of "$B(h)$ should be p_2", and it is indeed the result of the process, no matter what p_1 is.

This distinction between "revision" and "updating" should not be confused with some other distinctions made in previous discussions. For example, some authors use "revision" for "correction of wrong beliefs", and "updating" for "belief change due to changing world" [Dubois and Prade (1991)]. In the current discussion, however, the issue is not why a belief change happens, but whether the new value is exactly what the evidence says.

Another distinction is "revision" vs. "focusing", as discussed in Dubois and Prade (1997), corresponding to the modification of generic knowledge and the shifting of reference class according to the current situation, respectively. Focusing does not really change any belief of the system. Instead, in this process conditional propositions are used to replace unconditional propositions (though the conditions may not be explicitly represented). On the other hand, the revision view of conditioning, as advocated by people working in probability kinematics [Jeffrey (1965); Domotor (1980)], is about how to change a probability distribution according to new information. As analyzed previously, conditioning used in this way can only completely replace the previous probability distribution with a new one (what I call "updating"), but cannot balance (or take some kind of "average" of) the two (what I call "revision").

Though "updating" is a valid operation in uncertain reasoning, it is only a special case of "revision", because it does not cover the situation where the result is a compromise of conflicting beliefs/information/evidence. In certain situations, it is proper to interpret belief changes as updating [Dubois and Prade (1991)], but revision is a more general and important operation. When there are conflicts among beliefs, it is unusual that one piece of evidence can be *completely* suppressed by another piece of evidence, even though it make sense to assume that new evidence is usually "stronger" than old evidence.

Some authors represent revision as "deriving $P(h|e_1 \wedge e_2)$ from $P(h|e_1)$ and $P(h|e_2)$" [Deutsch-McLeish (1991)]. According to the previous discussion, we can see that this treatment only considers explicit conditions, while in general we cannot assume that conflict beliefs always come under the same implicit condition.

Concretely speaking, revision of probability happens when the system's current belief on h is $P_{K_0}(h)$, the new knowledge is $P_{K'_0}(h)$, and the result is $P_{K_1}(h)$, where K_1 summarized the knowledge in K_0 and K'_0. We cannot do it in the Bayesian approach, because $P_{K'_0}(h)$ contains information that cannot be derived from P_{K_0}, nor can the operation be treated as updating, where $P_{K_0}(e)$ is simply replaced by $P_{K'_0}(h)$. Intuitively, to carry out the revision operation, we need more information about K_0 and K'_0, and this information is not in the probability distribution function P_{K_0}.

Therefore, even if Jeffrey's rule is used to replace Bayes' Theorem and structure information is added into the picture, the system still does not have a general way to revise its implicit conditions (i.e., background knowledge behind the probability distribution function). If we want to apply a Bayesian network to a practical domain, one of the following requirements must be satisfied:

(1) The implicit condition of the initial probability distribution, that is, the domain knowledge used to determine the distribution initially, can be assumed to be immune from future modifications; or
(2) All modifications of the implicit condition can be treated as updating, in the sense that when new knowledge conflict with old knowledge, the latter is completely abandoned.

From artificial intelligence's point of view, such domains are exceptions, rather than general situations. In most cases, we can guarantee neither that all initial knowledge is unchangeable, nor that later acquired knowledge always completely suppresses earlier acquired knowledge. Usually, revision is a compromise, as addressed in the discussions on belief change [Voorbraak (1999)] and multiple source information fusion [Dubois et al. (2001)]. [3]

Some people may think that whenever the above situation happens, we can always go back to the very beginning, to redefine the proposition space and the probability function on it, according to all currently available information. Of

[3] Information fusion also includes many other issues, such as to detect whether the information sources are using the same evidence. These issues are beyond the scope of this paper.

course this can be done, but it is not done by the Bayesian learning mechanism of the system itself, but by the designers using something else. Such a redesign shows the limitation, not the strength, of Bayesian learning.

4 Ignorance and sensitivity

Since the probability distribution function P is defined on S according to implicit condition K, it provides a summary of the available knowledge of the system about propositions in S, but the function says little about K itself. Consequently, the system has no general way to revise and extend K. The Bayesian approach has no general way to represent and handle the uncertainty within the background knowledge and the prior probability function. This is a serious limitation of Bayesianism, both in theory and in application.

Though the distinction between explicit and implicit conditions is rarely made, the above conclusion, that is, Bayesianism has limitations in representing and processing uncertainty, is not new at all. From different considerations, many people reached the same conclusion, that is, to use a probability distribution function alone to represent uncertainty is not enough, because it fails to show the *ignorance*, or uncertainty about the function itself.

Several alternative approaches are proposed to solve this problem, including the following:

- probability interval, where the probability value is not specified as a point, but as an interval, and the width of the interval indicates the ignorance of the system [Grosof (1986); Kyburg (1988)],
- higher-order probability, where a second probability value is introduced to specify the accuracy of the "first-order" probability evaluation [Kyburg (1988); Paaß (1991)],
- imprecise probability, where upper and lower probability values are used to replace precise probability values [Walley (1991, 1996)],
- Dempster-Shafer theory, where a belief function and a plausibility function are used to represent uncertainty, and a evidence combination rule is used to reduce ignorance [Dempster (1967); Shafer (1976)],
- confidence measurement, where a frequency value and a confidence value are used to represent uncertainty, and the latter also indicates ignorance [Wang (1993, 2001)].

Though these approaches are technically different, they can all (more or less) be seen as attempts of extending the Bayesian approach by using more than one value to represent the uncertainty of a statement, and therefore indicates the ignorance of the system. [4]

To discuss these approaches is beyond the scope of this research note. [5] In the following I only analyze the response from the Bayesian school against these challenges.

To argue against the opinion that "more than one number is needed to represent uncertainty", Cheeseman (1985) claimed that a point value and a density function will give the same result in decision making, which I agree to certain extent. However, I believe that he was wrong by saying that standard deviation can be used to capture "the change of expectations" (or revision, as defined in this paper). If we test a proposition n times, and the results are the same, then the standard deviation of the results is 0, that is, independent to n. But our confidence about "the result will remain the same" will obviously increase with n. Actually, what the standard deviation measures is the *variations* among the samples, but what needed in revision, intuitively speaking, has more to do with the *amount* of the samples.

Pearl said the uncertainty in the assessment of $P_0(e)$ is measured by the (narrowness of the) distribution of $P_0(e|c)$ as c ranges over all combinations of contingencies, and each combination c is weighted by its current belief $P_0(c)$ [Pearl (1988)]. A similar approach is in Spiegelhalter (1989), where ignorance is treated as sensitivity.

I agree with them that ignorance is the lack of confidence, and confidence can be measured by how much a degree of belief can be modified by possible future evidence (in this sense, it is different from what measured by the "confidence interval" in statistics) [Wang (2001)]. However, in their definition, they still assume that all relevant future evidence causing a belief change can be represented as an *explicit condition*, and can be processed through conditioning. As a result, their measurement of ignorance (or confidence) cannot captures the ignorance about *implicit conditions*.

No matter whether other approaches can solve the problem, as far as the "ignorance" to be represented is about an implicit condition, it cannot be handled properly by Bayesianism. For a specific domain, if revision is a crucial operation for the solving of the practical problems, the Bayesian approach cannot be used, and other approaches should be considered.

[4] A related work is possibility theory, which is not an attempt to extend probability theory, but also use two values to represent ignorance [Dubois et al. (1996)]. There are other competitors of Bayesianism, such as fuzzy logic [Zadeh (1965)] and Spohn's kappa calculus [Spohn (1990)], as well as the approaches without numerical measurement of uncertainty, but they are irrelevant to the current discussion.

[5] Such discussions can be found in Wang (1994, 1995, 2001).

5 Conclusion

In this short research note I do not intend to evaluate all aspects of Bayesianism, not even all of its limitations. [6] Instead, I try to clarify the interpretation of conditional probability by distinguish two types of conditions, because this distinction is in the center of several previous debates, and the confusion can be found in many influential publications.

According to the above analysis, what the Bayesian approach can do is:

- Given some values in a probability distribution, to calculate some other values in the same distribution, as shown in equation (1).
- Given some values in a new probability distribution, to update a previous probability distribution accordingly, as shown in equations (2) and (5).

What it cannot do is:

- To combine conflicting beliefs that are based on different implicit conditions, such as $P_{K_0}(h)$ and $P_{K'_0}(h)$.
- To carry out inference when the premises are based different implicit conditions, such as $P_{K_0}(h|e)$ and $P_{K'_0}(e)$.

For the last two cases, we need additional information about the implicit conditions involved (to merely attach a probability value to each belief is not enough), as well as inference rules that use this kind of information.

When a reasoning system has insufficient knowledge and resources (with respect to the tasks assigned to it), it cannot assume that the initial background knowledge does not need revise, nor that all revisions can be treated as complete updating of probability distribution function. Therefore, the above limitation means that the Bayesian approach is not a normative theory of reasoning in this situation, and we need something else [Wang (2001)].

Though similar conclusions were proposed by other people before, the discussion has been messed up by the confusion between explicit conditions and implicit conditions of probability evaluations. This confusion is both conceptual and notational, and it causes a serious underestimation about the limitation of Bayesianism. To clearly distinguish these two types of conditions, as well as to clearly distinguish different operations like revision and updating, will not only help us to understand the capacity of the Bayesian approach, but will also help us to design and analyze alternative approaches for reasoning under uncertainty.

---

[6] There are other issues, such as the requirement of a coherent prior probability function and the computational expense of global updating.

Acknowledgment

Thanks to the anonymous reviewers for helpful comments.

References

Cheeseman, P., 1985. In defense of probability. In: Proceedings of the Eighth International Joint Conference on Artificial Intelligence. pp. 1002–1009.

Cheeseman, P., 1986. Probabilistic versus fuzzy reasoning. In: Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam, pp. 85–102.

Coletti, G., Gilio, A., Scozzafava, R., 1993. Comparative probability for conditional events: a new look through coherence. Theory and Decision 35, 237–258.

Dempster, A., 1967. Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339.

Deutsch-McLeish, M., 1991. A study of probabilities and belief functions under conflicting evidence: Comparisons and new methods. In: Bouchon-Meunier, B., Yager, R. R., Zadeh, L. A. (Eds.), Uncertainty in Knowledge Bases: Proc. of the 3rd International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'90. Springer, Berlin, Heidelberg, pp. 41–49.

Diaconis, P., Zabell, S., 1983. Some alternatives to Bayes's rule. In: Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine, Conference on Political Economy. pp. 25–38.

Domotor, Z., 1980. Probability kinematics and representation of belief change. Philosophy of Science 47, 384–403.

Dubois, D., Grabisch, M., Prade, H., Smets, P., 2001. Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: The assessment of the value of a candidate. International Journal of Intelligent Systems 16, 1245–1272.

Dubois, D., Prade, H., 1991. Updating with belief functions, ordinal conditional functions and possibility measures. In: Bonissone, P., Henrion, M., Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence 6. North-Holland, Amsterdam, pp. 311–329.

Dubois, D., Prade, H., 1997. Focusing vs. belief revision: A fundamental distinction when dealing with generic knowledge. In: Gabbay, D. M., Kruse, R., Nonnengart, A., Ohlbach, H. J. (Eds.), Proceedings of the First International Joint Conference on Qualitative and Quantitative Practical Reasoning. pp. 96–107.

Dubois, D., Prade, H., Smets, P., 1996. Representing partial ignorance. IEEE Trans. on System, Man, and Cybernetics 26, 361–377.

Grosof, B., 1986. An inequality paradigm for probabilistic knowledge. In: Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam, pp. 259–275.

Heckerman, D., 1988. An axiomatic framework for belief updates. In: Lemmer, J., Kanal, N. (Eds.), Uncertainty in Artificial Intelligence 2. North-Holland, Amsterdam, pp. 11–22.

Heckerman, D., 1999. Bayesian learning. In: Wilson, R., Keil, F. (Eds.), The MIT Encyclopedia of the Cognitive Sciences. MIT Press, Cambridge, Massachusetts, pp. 70–72.

Jeffrey, R., 1965. The Logic of Decision. McGraw-Hill, New York.

Kolmogorov, A. N., 1950. Foundations of the Theory of Probability. Chelsea Publishing Company, New York.

Kyburg, H., 1987. Bayesian and non-Bayesian evidential updating. Artificial Intelligence 31, 271–293.

Kyburg, H., 1988. Higher order probabilities and intervals. International Journal of Approximate Reasoning 2, 195–209.

Paaß, G., 1991. Second order probabilities for uncertain and conflicting evidence. In: Bonissone, P., Henrion, M., Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence 6. North-Holland, Amsterdam, pp. 447–456.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers, San Mateo, California.

Pearl, J., 1990. Jeffrey's rule, passage of experience, and Neo-Bayesianism. In: Kyburg, H., R., L., G., C. (Eds.), Knowledge Representation and Defeasible Reasoning. Kluwer Academic Publishers, Amsterdam, pp. 245–265.

Pearl, J., 2000. Causality. Cambridge University Press, Cambridge, UK.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey.

Spiegelhalter, D., 1989. A unified approach to imprecision and sensitivity of beliefs in expert systems. In: Kanal, L., Levitt, T., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence 3. North-Holland, Amsterdam, pp. 199–208.

Spohn, W., 1990. A general non-probabilistic theory of inductive reasoning. In: Shachter, R., Levitt, T., Kanal, L., Lemmer, J. (Eds.), Uncertainty in Artificial Intelligence 4. North-Holland, Amsterdam, pp. 149–159.

Voorbraak, F., 1999. Probabilistic belief change: Expansion, conditioning and constraining. In: Uncertainty in Artificial Intelligence: Proceedings of the Fifteenth Conference (UAI-1999). Morgan Kaufmann Publishers, San Francisco, CA, pp. 655–662.

Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London.

Walley, P., 1996. Inferences from multinomial data: learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58, 3–57.

Wang, P., 1993. Belief revision in probability theory. In: Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence. Morgan Kauf-

mann Publishers, San Mateo, California, pp. 519–526.
Wang, P., 1994. A defect in Dempster-Shafer theory. In: Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Mateo, California, pp. 560–566.
Wang, P., 1995. An unified treatment of uncertainties. In: Proceedings of the Fourth International Conference for Young Computer Scientists. Beijing, pp. 462–467.
Wang, P., 2001. Confidence as higher-order uncertainty. In: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications. Ithaca, New York, pp. 352–361.
Zadeh, L., 1965. Fuzzy sets. Information and Control 8, 338–353.