A new look at conditional probability with belief functions

Ronald Meester | Timber Kerkvliet

Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Ronald Meester, Department of Mathematics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands.
Email: r.w.j.meester@vu.nl

We discuss repeatable experiments about which various agents may have different information. This information can vary from a full probabilistic description of the experiment in the sense that the probabilities of all outcomes are known to the agent, to having no information whatsoever, except the collection of possible outcomes. We argue that belief functions are very suitable for modeling the type of information we have in mind. We redevelop and rederive various notions of conditional belief functions, using a viewpoint of relative frequencies. We call the two main forms of conditioning contingent and necessary conditioning, respectively. The former is used when the conditioning event may also have not occurred, whereas the latter is used when it turns out that the event on which we condition occurs necessarily. Our approach unifies various notions in the literature into one conceptual framework, namely, the updated belief functions of Fagin and Halpern, the unnormalized conditional belief function of Smets, and the notions of updating and focusing as used by Dubois and Prade. We show that the original Dempster–Shafer definition of conditional belief functions cannot be interpreted directly in our framework. We give a number of examples illustrating our interpretation, as well as the differences between the various notions of conditioning.

KEYWORDS
belief function, conditional probability, contingent conditioning, necessary conditioning

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Statistica Neerlandica published by John Wiley & Sons Ltd on behalf of VVS.
1 | MODELING PARTIAL INFORMATION

In this article, we are concerned with (in principle) repeatable experiments about which various agents may have different information. This information can vary from a full probabilistic description of the experiment in the sense that the probabilities of all outcomes are known to the agent, to having no information whatsoever, except the collection of possible outcomes. Between these two extremes, we have situations in which agents have partial information in a specific form. To illustrate what kind of situations we have in mind, here is a first introductory example.

Example 1. Suppose that an experiment results in a number between 1 and 10, inclusive. If the actual choice is uniformly at random, then a fully informed agent will assign probability 1/10 to each of the 10 possible outcomes.

However, an agent who is only told that the probability of an even number is 1/2 and that the probability of an odd number is also 1/2 has partial information about the experiment. This agent would assign probability 1/2 to the set \{1, 3, 5, 7, 9\} and probability 1/2 to the set \{2, 4, 6, 8, 10\}.

The most extreme situation is the one in which an agent receives no information whatsoever about how the choice of the number is made. For this agent, the only meaningful thing is to assign probability 1 to the full set of outcomes 1, 2, …, 10.

How would each of the agents in Example 1 assess the probability that the experiment results in a prime? When we think of probabilities as describing relative frequencies, then the first agent knows that when the experiment is repeated many times, roughly 4 out of 10 times the outcome will be prime.

As far as the second agent is concerned, for all he knows, the actual experiment could possibly consist of drawing a 9 with probability 1/2, and drawing a 6 with probability 1/2. Such an experiment would be consistent with his knowledge but would never result in a prime. Obviously, the third agent knows even less.

Therefore, depending on the information of the various agents, they draw different conclusions about the event of seeing a prime. If they were to repeatedly buy a bet on the event that the result is prime, and the bet would pay out 1 if the actual result is in fact prime, then the first agent may reasonably be willing to pay up to 4/10 for such a bet. For the second and third agents, however, it would not be an unreasonable position to decide not to pay anything at all for such a bet. Indeed, their knowledge is consistent with an experiment in which a prime never shows up and betting any positive amount could lead to losing a lot of money when done repeatedly. In other words, their knowledge is incomplete to the effect that a betting a prime might lead to a sure loss. Although of course not forced, this means that not betting is a reasonable position. Here is a second example.

Example 2. Consider a situation in which we have two coins and a die. Someone rolls the die and follows the following instruction. If he rolls 1 or 2, he puts both coins with heads facing up. If he rolls 3 or 4, he puts both coins with tails facing up. If he rolls 5 or 6, we have no information about what the agent does. In particular, we do not even know whether in this case the outcome can be described by a probability distribution at all.

How can we describe our partial knowledge of this experiment? Let

\[ \Omega = \{(h, h), (h, t), (t, h), (t, t)\}, \]

where the first and second coordinates represent respectively the side of the first and second coins facing up (where “h” stands for heads and “t” stands for tails). The information we
MEESTER AND KERKVLIET

have can be summarized by assigning probability 1/3 to each of the sets \{(h, h)\}, \{(t, t)\}, and \(\Omega\). This assignment precisely captures the idea that we know that with probability 1/3, two heads will be up; with probability 1/3, two tails will be up; and with probability 1/3, we have no information whatsoever.

Note that we do not assume that the process that leads to the position of the coins can be completely described by some underlying probability measure. We have no idea what happens when the die shows 5 or 6, and for all we know, the outcome may be deterministic in that case or may depend on nonrandom things we are not aware of. Assigning mass 1/3 to \(\Omega\) simply represents total ignorance about what happens when the die shows 5 or 6.

Situations as described in Examples 1 and 2 can be conveniently modeled with belief functions which we introduce now. The following definitions go back to Shafer (1976). We start with a mapping on the subsets of a finite outcomespace \(\Omega\).

**Definition 1.** A function \(m : 2^\Omega \to [0, 1]\) is called a basic belief assignment if \(m(\emptyset) = 0\) and
\[
\sum_{C \subseteq \Omega} m(C) = 1. \tag{2}
\]
The mapping \(m\) naturally defines a probability measure \(P\) on the space \(2^\Omega\) via
\[
P(C) = \sum_{C \in C} m(C),
\]
for any collection \(C\) of subsets of \(\Omega\). Note that \(P(\{C\}) = m(C)\) for all \(C \subseteq \Omega\).

Shafer defines the belief in \(A \subset \Omega\) as the probability that when we choose a subset according to \(m\), the outcome implies \(A\):

**Definition 2.** The function Bel : \(2^\Omega \to [0, 1]\) given by
\[
\text{Bel}(A) = \sum_{C \subseteq A} m(C) \tag{3}
\]
is called the belief function corresponding to the basic belief assignment \(m\).

It is well known (see Shafer, 1976) for a proof) that \(m\) can be retrieved from Bel via
\[
m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \text{Bel}(B),
\]
where \(|\cdot|\) denotes cardinality. Note that Bel is a probability distribution if and only if \(m\) concentrates on atomic events.

**Definition 3.** The plausibility of an event \(A\) is defined as
\[
\text{Pl}(A) := 1 - \text{Bel}(A^c) = \sum_{S \cap A \neq \emptyset} m(S). \tag{4}
\]

The connection between these definitions and the examples above is as follows. Each agent in Example 1 has its own \(m\), and therefore, also his own belief function. The \(m\) corresponding to the first agent is just the uniform probability measure on the outcome space. The \(m\) of the second agent is given by \(m((1, 3, 5, 7, 9)) = m((2, 4, 6, 8, 10)) = 1/2\) and \(m(A) = 0\) for all other subsets of the outcome space. The third agent corresponds to \(m\), which puts all probability mass on the set \(\Omega\) itself. This latter belief assignment corresponds to having no information apart from the relevant outcome space.
How should we now interpret Bel(A) and Pl(A)? One interpretation of the belief in A is the probability that A is implied by the outcome of an experiment. This interpretation goes back to Pearl (1990), who calls this “the probability of provability.” An application where this interpretation is used is Kerkvliet and Meester (2016), in a forensic context.

Dubois and Prade (1992) offer, among others, the following interpretation. (We briefly comment on their other interpretations in the last section.) According to them, any random experiment whose outcomes cannot be precisely observed lends itself to a treatment by belief functions. They see m(S) as the frequency of making observation S when the experiment is repeated, meaning the frequency that we learn precisely that the truth is in S, and nothing else.

This idea can be formalized as follows. Consider a random vector (X, Y), where X takes values in Ω and Y takes values in the subsets of Ω. The joint distribution Q of (X, Y) must be so that Q(X ∈ Y) = 1. We assume that we know the marginal distribution of Y, namely, the one given by m. What can we say about the marginal distribution of X? Clearly, the fact that we assume X ∈ Y implies that

\[ Q(X ∈ A) ≥ \sum_{B ⊆ A} Q(Y = B) \quad \text{and} \quad Q(X ∈ A) ≤ 1 - \sum_{B \cap A = \emptyset} Q(Y = B) \quad (5) \]

so that, writing Bel for the belief function corresponding to m, we have

\[ \text{Bel}(A) ≤ Q(X ∈ A) ≤ \text{Pl}(A). \quad (6) \]

In Fagin and Halpern (1989), the collection of all probability distributions Q on Ω for which Bel(A) ≤ Q(A) ≤ Pl(A) for all A ⊆ Ω is called the collection of consistent probability distributions. The belief in an event A in this context boils down to the minimum probability of A within that collection (called a lower probability or a lower envelope). Indeed, it is not hard to see that for any A ⊆ Ω, there is a consistent probability measure for which the lower bound in (6) is an equality and similarly for the upper bound. A useful way to think about the relation between the basic belief assignment m and the consistent probability distributions is to think of m(A) as the probability mass that is confined to A, which however can be distributed over the elements of A in any desired way.

Although the position in Dubois and Prade (1992) and Fagin and Halpern (1989) is elegant and useful, we prefer not to make the assumption that the pair (X, Y) has a joint distribution. We prefer to only assume that there is an outcome (x, y) of the experiment, with x ∈ y and such that the second coordinate is described by a marginal distribution function determined by m. This really corresponds to having no information whatsoever about x other than that it is contained in y. Within this interpretation, Bel(A) can be interpreted in terms of a statement concerning relative frequencies, as we see below. We remark, though, that within the framework of consistent probability distributions that we described above, versions of all our theorems can also be formulated and proved, in very much the same way as we do.

Suppose we repeat the experiment many times, that is, we have (x_1, C_1), (x_2, C_2), …, where x_i ∈ C_i for all i and where the distribution of the C_i is known and described by m. The C_i are observed. The x_i are not observed and we only know that x_i ∈ C_i for all i. For C = (C_1, C_2, C_3, …), we write

\[ \mathcal{X}(C) := \{ (x_1, x_2, x_3, …) : x_i ∈ C_i \} \quad (7) \]

for the collection of all realizations compatible with observing C. We write \( P^∞ \) for the infinite-dimensional product probability distribution capturing independent drawings of subsets of Ω according to P.
The following result expresses \( \text{Bel}(A) \) in terms of relative frequencies. Although the \( x_i \) are not observed, we have that the minimum limit inferior of the relative frequencies of \( A \) over all realizations compatible with our observations is precisely \( \text{Bel}(A) \). This may seem a somewhat overly complicated way to characterize \( \text{Bel}(A) \), but by using this formula, we prepare for similar expressions concerning conditional beliefs in the next section.

\[ \text{Theorem 1. For all } A \subseteq \Omega, \text{ we have} \]

\[ P^\infty \left( C, \inf_{x \in \mathcal{X}(C)} \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i) \right) = \text{Bel}(A) = 1. \]  

Moreover, the infimum is attained \( P^\infty \), almost surely.

**Proof.** Let \( C = (C_1, C_2, C_3, \ldots) \) be a sequence of subsets of \( \Omega \) and let \( y \in \mathcal{X}(C) \) be chosen such that \( y_i \in A \) if \( C_i \subseteq A \) and \( y_i \in A^c \) if \( C_i \cap A^c \neq \emptyset \). Clearly, we have

\[ \frac{1}{n} \sum_{i=1}^{n} 1_A(y_i) \leq \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i) \]  

for all \( x \in \mathcal{X}(C) \) and all \( n \). Hence,

\[ \inf_{x \in \mathcal{X}(C)} \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(y_i) \]

\[ = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} I_i, \]  

where \( I_i = 1 \) if \( C_i \subseteq A \) and \( I_i = 0 \), otherwise. The strong law of large numbers tells as that the last \( \liminf \) is in fact almost surely a limit and equal to

\[ \sum_{S \subseteq A} m(S) = \text{Bel}(A). \]

The type of information that can be described with belief functions in our interpretation seems rather natural and intuitive. The first goal of the present paper is to convince the reader that belief functions can be successfully used to model partial information of agents in many situations. To this end, we give more examples below and in Section 4.

The second and main goal of this paper is to introduce a new way of looking at conditional beliefs and to show that this new angle leads to satisfactory and understandable results in applications. In the literature, a number of suggestions for conditional (sometimes also called “updated”) beliefs are developed. In his original book, Shafer (1976) arrived at a notion of conditioning as a special case of Dempster’s rule of combination. In Fagin and Halpern (1989), another notion of conditional belief was developed, based on the interpretation that belief function express uncertainty about the underlying classical probability measure of an experiment. Dubois and Prade (1992) distinguish between “focusing” and “updating” and suggest different formulas for either type of conditioning; see also Jaffray (1990). Finally, there is Smets’s unnormalized conditioning (see, e.g., Cuzzolin, 2016) developed in yet another context.

The main contribution of this paper is, in brief, as follows. We approach an conditional belief from frequentistic reasoning on the level of the basic belief assignment, and we place the various notions of conditional belief into one conceptual framework. To this end, we make a new caesura and distinguish between contingent and necessary conditioning. When an agent receives information upon which he wants to update his beliefs, it matters whether or not the information is
about the outcome of an experiment or about the structure of the experiment itself. The former case leads to contingent conditioning, whereas the latter case leads to necessary conditioning.

Although we develop the various notions of conditioning independently, it turns out that contingent conditioning leads to the same formula as the updated belief functions in Fagin and Halpern (1989) and the “focusing” version of updating belief functions in Dubois and Prade (1992). The notion of necessary conditioning turns out to lead to a formula that is a special case of Smets’s unnormalized conditioning. In our interpretation, there seems to be no place for the original Dempster–Shafer (DS) notion of conditional beliefs as developed in Shafer (1976), and we will explain why this is so.

The next well-known example shows that the distinction between necessary and contingent conditioning is meaningful.

Example 3 (The Monty Hall problem).
Consider three boxes, numbered 1, 2, and 3. One box contains a prize; the other two are empty. The candidate chooses one of the boxes but does not open it. Monty Hall subsequently opens another box that is empty, and the candidate is offered the possibility to swap to the third (unopened) box. Should he do this? We can model this with a belief function as follows. Assume that the box that contains the prize is chosen according to a classical probability distribution so that box \( i \) is chosen with probability \( p(i) \), where \( p(1) + p(2) + p(3) = 1 \). Further suppose that the initial choice of the candidate is uniformly distributed; he simply makes a random guess. Set \( \Omega = X^3 \) where \( X = \{1, 2, 3\} \) and let \( P \) denote the box containing the prize, \( S \) the box initially chosen, and \( O \) the box opened by Monty Hall. The coordinates of \( \omega = (\omega_1, \omega_2, \omega_3) \in \Omega \) correspond to \( P, S, \) and \( O \), in that order.

Now, we set the basic belief assignment from the viewpoint of the candidate, as follows:

\[
m(\{x\} \times \{y\} \times X \setminus \{y\}) = \frac{1}{3} p(x),
\]

for all \( x, y \in X \). Indeed, the position of the prize and the choice of the candidate are independent, and all we know at this point is that Monty Hall will open a box that was not chosen by the candidate. It is now easy to compute that \( \text{Bel}(P = S) = 1/3 \), as it should.

Next, we want to incorporate the fact that Monty Hall opens a box that does not contain the prize. Hence, we want to condition on the event \( H = \{O \neq P\} \), and we are interested in the conditional belief in the event that \( S = P \), conditioned on \( H \). The key question now is whether or not Monty Hall knows the position of the prize and hence necessarily opens an empty box or that he is ignorant and happens to open an empty box. We will see below that the difference is significant and leads to different answers. Necessary conditioning leads to the classical answer, but contingent conditioning does not.

At the end of this introduction, we position our article in the context of epistemic interpretations of probability, that is, uncertainty having to do with partial knowledge rather than with objective facts. Epistemic probability cannot always be satisfactorily described with the classical axioms of Kolmogorov, something which has been recognized and confirmed by many researchers from such different disciplines as mathematics (Cohen, 1977; Marinacci, 1999; Royall, 1997; Shafer, 1976, 1981, 2008; Walley, 1991; Walley & Fine, 1982), legal science (Cohen, 1977), and philosophy (see Dietz, 2010; Fagin & Halpern, 1989; Haenni, 2009; Walley, 1991, and the references therein). These authors have argued that the classical axioms of probability are too restrictive, for various reasons. From Example 1, we already see that rational betting behavior cannot always be captured by a classical probability measure. Indeed, the belief functions of the second and third agents are no probability measures.
In the literature, several arguments are given why epistemic probability does not always comply to the usual axioms. To name just two of them: (1) It is impossible for an agent to distinguish between disbelief in \( A \), by which we mean \( P(A^c) \), and lack of belief, by which we mean \( 1 - P(A) \). Lack of supporting evidence for \( A \) should not necessarily be supporting evidence for \( A^c \); (2) It is impossible to model complete ignorance with classical probability. Often, one resorts to uniform prior probabilities, but this habit is not well founded (see, e.g., Royall, 1997; Shafer, 1976). Ignorance cannot be modeled with a uniform distribution.

The belief functions as originally developed by Shafer (1976) are one out of several attempts to develop a notion of probability or belief, which better suits the needs for epistemic interpretations of probability. Other approaches can be found in Cohen (1977) and Walley (1991). The approach of Walley (1991) is in fact related to belief functions, as follows: We interpret belief functions as the amount of money an agent wishes to maximally pay for a bet on an event \( A \). In Walley’s formulation, these betting amounts are formalized into his lower provisions. We have shown elsewhere (Kerkvliet & Meester, 2016) that a belief function is in fact characterized as a lower provision with one extra weak requirement of consistency.

In the next section, we develop contingent conditioning. In Section 3, we develop necessary conditioning, and in Section 4, we extensively discuss two examples illustrating the difference between classical, contingent, and necessary conditioning. After that, we explain how the classing DS conditioning relates to our approach, and we end with a concluding section in which we relate the various forms of conditioning to the existing notions mentioned above.

## 2 | CONTINGENT CONDITIONING

In classical probability theory, one of the most convincing ways to think about conditional probability is to consider many repetitions of the same experiment. Conditioning on an event, \( H \) then means that we simply ignore all outcomes that are not in \( H \), and the conditional probability of an event \( A \) given \( H \) is then given by the relative frequency of \( A \) in the resulting subsequence.

This idea can and has been (Dubois & Prade, 1992; Fagin & Halpern, 1989) used in the context of belief functions as well. The problem is, however, that we do not always know whether or not \( A \) and/or \( H \) actually occurs because we do not receive complete information about the outcome of the experiment. Indeed, when we receive information that the truth is in \( S \), or that \( S \) occurs, this may be compatible with both \( H \) and \( H^c \) and similarly for \( A \).

However, we can, in the same spirit as Theorem 1, compute the largest lower bound of the limit inferior of the relative frequency of and event \( A \) relative to \( H \), given the information we have. This then naturally leads to a definition of conditional belief.

**Theorem 2.** For all \( A \subseteq \Omega \), we have

\[
P^\infty \left( C; \inf_{x \in \mathcal{C}(C)} \left( \liminf_{n \to \infty} \frac{\sum_{i=1}^{n} 1_{A \cap (H^c \cup A^c)}(x_i)}{\sum_{i=1}^{n} 1_{H^c}(x_i)} \right) \right) = \frac{\text{Bel}(A \cap H)}{\text{Bel}(A \cap H) + \text{Pl}(H^c \setminus A)} = 1. \tag{11}
\]

Moreover, the infimum is attained \( P^\infty \), almost surely.

**Proof.** Let

\[ C_1 := \{ C \subseteq \Omega : C \subseteq A \cap H \} \]

and

\[ C_2 := \{ C \subseteq \Omega : C \cap (H \setminus A) \neq \emptyset \}. \]
Choose \( y \in \mathcal{A}(C) \) in such that

1. \( y_i \in H \cap A \), if \( C_i \in C_1 \),
2. \( y_i \in H \setminus A \), if \( C_i \in C_2 \),
3. \( y_i \in H \setminus F \), if \( C_i \not\in C_1 \cup C_2 \).

We claim that

\[
\frac{\sum_{i=1}^n 1_{A \cap H}(x_i)}{\sum_{i=1}^n 1_H(x_i)} \geq \frac{\sum_{i=1}^n 1_{A \cap H}(y_i)}{\sum_{i=1}^n 1_H(y_i)}.
\] (12)

for all \( x \in \mathcal{A}(C) \) and for all \( n \) for which the denominators are nonzero.

To see this, we observe that (12) is true if and only if

\[
\frac{\sum_{i=1}^n 1_{A \cap H}(x_i)}{\sum_{i=1}^n 1_{A \cap H}(x_i) + \sum_{i=1}^n 1_{A \cap H}(x_i)} \geq \frac{\sum_{i=1}^n 1_{A \cap H}(y_i)}{\sum_{i=1}^n 1_{A \cap H}(y_i) + \sum_{i=1}^n 1_{A \cap H}(y_i)},
\] (13)

which is equivalent with

\[
\left( \sum_{i=1}^n 1_{A \cap H}(x_i) \right) \left( \sum_{i=1}^n 1_{A \cap H}(y_i) \right) \geq \left( \sum_{i=1}^n 1_{A \cap H}(y_i) \right) \left( \sum_{i=1}^n 1_{A \cap H}(x_i) \right).
\] (14)

By construction, we have \( 1_{A \cap H}(x_i) \geq 1_{A \cap H}(y_i) \) and \( 1_{A \cap H}(x_i) \leq 1_{A \cap H}(y_i) \), and thus, (14) holds. Hence,

\[
\inf_{x \in \mathcal{A}(C)} \liminf_{n \to \infty} \frac{\sum_{i=1}^n 1_{A \cap H}(x_i)}{\sum_{i=1}^n 1_H(x_i)} = \liminf_{n \to \infty} \frac{\sum_{i=1}^n 1_{A \cap H}(y_i)}{\sum_{i=1}^n 1_H(y_i)},
\]

which we can rewrite as

\[
\liminf_{n \to \infty} \frac{\frac{1}{n} \sum_{i=1}^n 1_{A \cap H}(y_i)}{\frac{1}{n} \sum_{i=1}^n 1_{A \cap H}(y_i) + \frac{1}{n} \sum_{i=1}^n 1_{H \setminus A}(y_i)}.
\]

It follows from the law of large numbers that \( \frac{1}{n} \sum_{i=1}^n 1_{A \cap H}(y_i) \) converges almost surely to \( \text{Bel}(A \cap H) \) and that \( \frac{1}{n} \sum_{i=1}^n 1_{H \setminus A}(y_i) \) converges to \( \text{Pl}(H \setminus A) \). Hence, the whole expression in fact converges almost surely, with limit

\[
\frac{\text{Bel}(A \cap H)}{\text{Bel}(A \cap H) + \text{Pl}(H \setminus A)}.
\]

This leads to the following definition.

**Definition 4** (Contingent conditioning).

Assume that \( \text{Pl}(H) > 0 \). We define the conditional belief in \( A \) given \( H \) by

\[
\text{Bel}_H(A) := \frac{\text{Bel}(A \cap H)}{\text{Bel}(A \cap H) + \text{Pl}(H \setminus A)}
\] (15)

if \( \text{Pl}(H \setminus A) > 0 \) and \( \text{Bel}_H(A) = 1 \) if \( \text{Pl}(H \setminus A) = 0 \). The corresponding conditional plausibility of \( A \) given \( H \) is given by

\[
\text{Pl}_H(A) = \frac{\text{Pl}(A \cap H)}{\text{Pl}(A \cap H) + \text{Bel}(H \setminus A)}.
\] (16)

We make four remarks.

1. One can also directly arrive at (16) by replacing the lim inf and inf by lim sup and sup, respectively, in Theorem 2. The proof of this completely parallels the proof of Theorem 2.
2. Note that if Bel is a probability distribution, we find
\[
\text{Bel}_H(A) = \text{Pl}_H(A) = \frac{\text{Bel}(A \cap H)}{\text{Bel}(H)} = \frac{\text{Pl}(A \cap H)}{\text{Pl}(H)},
\]
(17)
which means that this notion of conditioning generalizes the classical notion.

3. Formula (15) was obtained by Fagin and Halpern (1989) and by Dubois and Prade (1992). In their interpretation, a belief function describes the uncertainty we have about an underlying probability distribution. We, on the other hand, make no assumption of such underlying probability distribution, and our derivation is very different and more elementary than theirs.

4. In Fagin and Halpern (1989), it was shown that Bel_H is again a belief function if Bel(H) > 0. As such it corresponds in a one-to-one way to a basic belief assignment, which we denote by m_H.

We next give two examples. In the first example, we see that contingent conditioning is not commutative. The second example shows that the law of total probability does not hold. Both issues have been discussed in the literature (see, e.g., Pearl, 1990).

**Example 4.** Let \( \Omega = \{a, b, c, d\} \) and set
\[
m(\{a\}) = m(\{b, c\}) = m(\{c, d\}) = \frac{1}{3}.
\]
(18)
Suppose we condition on \( H = \{a, b, c\} \) in a contingent way. We find
\[
\text{Bel}_H(\{a\}) = \frac{\text{Bel}(\{a\})}{\text{Bel}(\{a\}) + \text{Pl}(\{b, c\})} = \frac{1}{3},
\]
(19)
and
\[
\text{Bel}_H(\{b, c\}) = \frac{\text{Bel}(\{b, c\})}{\text{Bel}(\{b, c\}) + \text{Pl}(\{a\})} = \frac{1}{2}.
\]
(20)
From these beliefs, we can reconstruct the corresponding conditional basic belief assignment:
\[
m_H(\{a\}) = \frac{1}{3}, \quad m_H(\{b, c\}) = \frac{1}{2}, \quad m_H(\{a, b, c\}) = \frac{1}{6}.
\]
(21)
Note that the belief in \( \{b, c\} \) is higher after conditioning, whereas the belief in \( \{a\} \) remained the same. Obviously, this cannot occur if Bel is a probability distribution: In that case, the ratios between the probabilities remain the same after conditioning. The reason that it happens in this example is that there is a chance of \( \frac{1}{3} \) on \( \{c, d\} \), which is consistent with both \( H \) and \( H^c \). Precisely for that reason, as described before, the beliefs in \( \{a\} \) and \( \{b, c\} \) are treated differently. For \( \{a\} \), the set \( \{c, d\} \) falls into category (3) and is counted in the computation. For \( \{b, c\} \) the set \( \{c, d\} \) falls into category (4) and is not counted.

Indeed, if we set \( B := \{a, b\} \), we find
\[
(\text{Bel}_B_H(\{a\}) = \frac{\text{Bel}_H(\{a\})}{\text{Bel}_H(\{a\}) + \text{Pl}_H(\{b\})} = \frac{1}{3},
\]
(22)
and
\[
(\text{Bel}_B_H(\{a\}) = \text{Bel}_B(\{a\}) = \frac{\text{Bel}(\{a\})}{\text{Bel}(\{a\}) + \text{Pl}(\{b\})} = \frac{1}{2}.
\]
(23)
If in general one wants to condition on two events \( H \) and \( J \), the most natural thing to do is to condition on \( H \cap J \) and not on \( H \) and \( J \) in any given order.
Example 5 (Example 2 continued).

The outcome space is given by

\[ \Omega = \{(h, h), (h, t), (t, h), (t, t)\}, \]

where the first and second coordinates represent respectively the side of the first and second coins facing up (here, “h” stands for heads and “t” stands for tails).

Let \( A = \{(t, t), (h, h)\} \) be the event that both coins have the same side facing up. It is now easy to check that \( \text{Bel}(A) = 2/3 \).

Now suppose that we are given the information that the second coin is heads. We write \( H = \{(h, h), (h, t)\} \) and compute

\[ \text{Bel}_H(A) = \frac{\text{Bel}((h, h))}{\text{Bel}((h, h)) + \text{Pl}((t, h))} = \frac{1}{2}. \]

In exactly the same way, we compute

\[ \text{Bel}_{H^c}(A) = \frac{\text{Bel}((t, t))}{\text{Bel}((t, t)) + \text{Pl}((h, t))} = \frac{1}{2}. \]

This might be somewhat surprising: If the chance is 1/2 in the case, we learn the second coin is heads as well as in the case we learn the second coin is tails, one might think that the unconditional chance should also be 1/2 because we know beforehand that the second coin is either heads or tails. However, we must be very careful with interpretation here. The quantity \( \text{Bel}_H(A) \) gives the belief in \( A \) contingent upon \( H \) and \( \text{Bel}_{H^c}(A) \) gives the belief in \( A \) contingent upon \( H^c \). While it is true that either \( H \) occurs or \( H^c \) occurs, it is clearly not true that either every time that \( H \) occurs the outcome is ignored or every time that \( H^c \) occurs the outcome is ignored. Therefore, we do not know at all beforehand that the belief in \( A \) should be considered either contingent upon the second coin being heads or contingent upon the second coin being tails. Hence, the law of total probability does not hold for contingent conditioning.

3 Necessary conditioning

Now, we consider updating according to the information that some set \( H \subseteq \Omega \) necessarily occurs. This must be understood in the context of the chance experiment: We learn that in every repetition of the experiment \( H \) occurs. This is information about the experiment rather than about an individual outcome. This means that this type of conditioning can only be used for updating if every outcome of the experiment \( C \subseteq \Omega \) that has positive probability (i.e., \( P(\{C\}) > 0 \)) is consistent with \( H \), that is, must satisfy \( C \cap H \neq \emptyset \). Hence, we must have \( \text{Pl}(H) = 1 \).

The viewpoint of the limit inferior of relative frequencies, in the same vein as in Theorems 1 and 2, is again useful here. To this end, for \( H \subseteq \Omega \) and \( C = \{C_1, C_2, C_3, \ldots\} \) as before, we set

\[ \mathcal{X}_H(C) := \{(x_1, x_2, x_3, \ldots) : x_i \in C_i \cap H\}. \]

That is, \( \mathcal{X}_H(C) \) is the collection of realizations compatible with our observation \( C \) and for which \( H \) always occurs. Note that \( \mathcal{X}_H(C) = \emptyset \) if \( C_i \subseteq H^c \) for some \( i \).

Theorem 3. Assume \( \text{Pl}(H) = 1 \). For all \( A \subseteq \Omega \), we have

\[ P^\infty\left(C; \inf_{x \in \mathcal{X}_H(C)} \left( \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i) \right) \right) = \text{Bel}(A \cup H^c) = 1. \]
Moreover, the infimum is attained $P^\infty$, almost surely.

Proof. Because $X_H(C) \neq \emptyset$, we can choose a $y \in X_H(C)$ such that $y_i \in A$ if $C_i \cap H \subseteq A$ and $y_i \in A^c$ if $C_i \cap H \cap A^c \neq \emptyset$. Then, it is the case that

$$\frac{1}{n} \sum_{i=1}^{n} 1_A(y_i) \leq \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i)$$

(29)

for all $x \in X_H(C)$ and for all $n$. Hence,

$$\inf_{x \in X_H(C)} \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(x_i) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} 1_A(y_i).$$

(30)

Because $y_i \in A$ precisely when $C_i \subseteq A \cup H^c$, the law of large numbers implies that the last limit inferior is in fact a limit almost surely and equal to $\text{Bel}(A \cup H^c)$.

This leads to the following definition.

**Definition 5** (Necessary conditioning). Assume that $\text{Pl}(H) = 1$. We define

$$\text{Bel}_{H, \text{necc}}(A) := \text{Bel}(A \cup H^c).$$

(31)

We make some remarks. First of all, our notion of necessary conditioning is very close to the notion of Smets’s unnormalized conditioning. Ours is technically more restrictive because we insist that $\text{Pl}(H) = 1$ when using it. In our frequentistic interpretation, we need this technical condition, as is clear also from the proof of Theorem 3.

Contrary to contingent conditioning, necessary conditioning is commutative, which seems to be what intuition requires. To show this, for $H, J \subseteq \Omega$ such that $\text{Pl}(H) = \text{Pl}(J) = 1$, we have

$$(\text{Bel}_{H, \text{necc}})_{J, \text{necc}}(A) = \text{Bel} ((A \cup H^c) \cup J^c) = \text{Bel} (A \cup (H \cap J)^c) = \text{Bel}_{H \cap J, \text{necc}}(A) = (\text{Bel}_{J, \text{necc}})_{H, \text{necc}}(A).$$

(32)

Similarly to contingent conditioning, necessary conditioning does not satisfy the law of total probability as the following example, in the spirit of Example 2, illustrates.

**Example 6.** We set $\Omega = \{(h, h), (h, t), (t, h), (t, t)\}$ as before, and define a basic belief assignment by

$$m (\{(h, h), (h, t)\}) = \frac{1}{2},$$

(33)

and

$$m (\{(t, h), (t, t)\}) = \frac{1}{2}.$$  

(34)

Let

$$E_0 = \{(h, h), (t, t)\}$$

(35)

be the event that the two coins have the same face showing up. We compute

$$\text{Bel}(E_0) = m (\{(h, h), (t, t)\}) + m (\{(h, h)\}) + m (\{(t, t)\}) = 0.$$

(36)
Next, we condition on the event $E_1 = \{(h, h), (t, h)\}$ that the second coin is heads. The sets \{(h, h), (h, t)\} and \{(t, h), (t, t)\} are the only outcomes with positive basic belief assignment and because they are both consistent with $E_1$, necessary conditioning applies. Intersected with $E_1$, they become respectively \{(h, h)\} and \{(t, h)\} so we find

$$\text{Bel}_{E_1, \text{ nec}}(E_0) = m(\{(h, h),(h, t)\}) = \frac{1}{2}.$$  \hspace{1cm} (37)

In exactly the same way, we find

$$\text{Bel}_{E_1, \text{ nec}}(E_0) = m(\{(t, h),(t, t)\}) = \frac{1}{2}.$$  \hspace{1cm} (38)

4 | COMPARING CONTINGENT AND NECESSARY CONDITIONING: TWO FURTHER EXTENSIVE EXAMPLES

In the following examples, we illustrate the difference between conditioning in the classical theory, contingent conditioning, and necessary conditioning. We also discuss under which circumstances any of these options should be applied.

Example 7. Consider $N + 1$ individuals, labeled $1, 2, \ldots, N + 1$ forming the space $X = \{1, 2, \ldots, N + 1\}$, an empty vase, and a fair coin. We flip the coin. If heads comes up, person number 1 puts a black ball in the vase while all other persons put a white ball in the vase. If tails comes up, then everybody puts a black ball in the vase.

Next, a ball is drawn from the vase, and we are told that the drawn ball is black. Is the drawn ball put in the vase by person number 1? As we will see below, our assessment of the answer to this question will depend on the (lack of) information we have about how the black ball was drawn. In particular, we cannot assume from the outset that the ball was chosen uniformly at random, and we should keep in mind that the draw of the ball may depend on the coin flip or on other things.

We work in the space

$$\Omega = \{0, 1\} \times \{B, W\}^{N+1} \times X.$$  

The first coordinate refers to the outcome of the coin flip, the next $N + 1$ coordinates refer to the color of the balls put in the vase by persons $1, 2, \ldots, N + 1$ (in that order) and the last coordinate refers to the origin of the drawn ball. We denote the coordinates by $A, \Gamma_1, \ldots, \Gamma_{N+1}$, and $F$, respectively.

Case 1 (Classical situation). Suppose that we are informed that the ball drawn from the vase is chosen randomly, with each ball being equally likely. We are now in a classical situation, which can be satisfactorily described with probabilities. To allow comparison with the upcoming cases, we however prefer to give the basic belief assignment $m_c$ corresponding to this case. This basic belief assignment is given by

$$m_c(0, B, B, \ldots, B, i) = \frac{1}{2} \frac{1}{N+1},$$  \hspace{1cm} (39)

and

$$m_c(1, B, W, \ldots, W, i) = \frac{1}{2} \frac{1}{N+1},$$  \hspace{1cm} (40)

for all $i = 1, \ldots, N + 1$. Because $m_c$ concentrates on atomic events, the corresponding belief function $\text{Bel}^c$ is a probability distribution.
Write $H = \{\Gamma_F = B\}$, the event that the drawn ball is black, and $E = \{F = 1\}$, the event that the drawn ball came from person 1. We are now interested in $\text{Bel}^c_H(E)$ (which is the same as the classical probability of $E$ given $H$ because $\text{Bel}^c$ is a probability distribution). We compute

$$
\text{Bel}^c_H(E) = \frac{\text{Bel}^c(E \cap H)}{\text{Bel}^c(E \cap H) + \text{Pl}^c(H \setminus E)}
= \frac{\frac{1}{N+1}}{\frac{1}{N+1} + \frac{\frac{1}{2}}{\frac{N}{N+1}}}
= \frac{2}{2 + N}.
$$

**Case 2** (NECESSARY CONDITIONING).

In Case 1, the procedure we had to model included the fact that the ball from the vase was randomly drawn. Without this information, the basic belief assignment $m$ we start out with has the following form:

$$m(A = 0, \Gamma_1 = B, \ldots, \Gamma_{N+1} = B) = \frac{1}{2},
$$

and

$$m(A = 1, \Gamma_1 = B, \Gamma_2 = W, \Gamma_3 = W, \ldots, \Gamma_{N+1} = W) = \frac{1}{2}.
$$

Indeed, in this case, we have no information whatsoever about the way the ball is drawn from the vase.

Now suppose that we do receive such information, namely, that the procedure is such that the drawn ball has to be black. This is, therefore, information that $H$ occurs necessarily, and we have to compute $\text{Bel}_{H,\text{neq}}(E)$. This is by definition equal to $\text{Bel}(E \cup IH)$, that is, the belief in the event that if $F \neq 1$, then $\Gamma_F = W$. It is easy to see that only the set in (42) implies this event, and therefore, the belief in this event is equal to $\frac{1}{2}$.

**Case 3** (CONTINGENT CONDITIONING).

Suppose now that we start out as in Case 2, that is, with the same basic belief assignment. This time however we do not get the information that the drawn ball is necessarily black, but simply learn that the outcome is a black ball. In this case we, therefore, need to condition in a contingent way. Starting out with the same belief function as in Case 2, it is easy to see that none of the sets in (41) and (42) is contained in $E \cap H$, and therefore, we find that $\text{Bel}_{H}(E) = 0$.

**Discussion.** We have distinguished between three cases, obtaining three different answers: $\frac{2}{2N+1}$, $\frac{1}{2}$, and 0. These are different answers that correspond to different situations, and we next argue that each of these answers is reasonable given the information we have.

In the first (classical case), we simply draw a uniform ball, and we asked what the probability is that the ball came from 1 given it was black. It is not a surprise that the answer tends to 0 when $N$ gets larger. Indeed, most instances where the drawn ball is black will come from a situation in which $A = 0$, in which case the probability that the ball came from 1 is only $1/(N+1)$.

In Case 2, we are sure that the drawn ball will be black. However, only when $A = 1$, are we sure that the drawn ball did come from 1. Indeed, when $A = 0$, everybody put a black ball in the vase. Because belief functions correspond to adding up the basic beliefs of sets which imply the target set, the answer $1/2$ is completely reasonable. Indeed, for all we know, it might be the case that only when $A = 1$ the black ball coming from 1 is chosen, and when $A = 0$, another black ball is chosen. There is no way we can rule out this scenario given the information we receive. Therefore, we cannot be sure that the relative frequency (when the
procedure would be repeated many times) will be larger than 1/2, and assigning conditional belief 1/2 to \( E \) is completely reasonable.

The situation in Case 3 is very different and can be compared with that in Case 1. Indeed, in both Cases 1 and 3, the concept of contingent conditioning is used. In Case 1, however, we have much more information about the way the procedure is carried out. In Case 3, we do not know how the balls are drawn, and we cannot exclude the possibility that the procedure is such that it will never be the case that a black ball coming from Person 1 is drawn. Hence, it is completely reasonable not to have any conditional belief in \( E \).

**Example 8** (The Monty Hall problem Example 3 continued).

We recall the situation. We assume that the probability that the prize is in box \( i \) is given by \( p(i) \) and that the initial choice of the candidate is uniformly distributed. Set \( \Omega = X^3 \), where \( X = \{1, 2, 3\} \) and let \( P \) denote the box containing the prize, \( S \) denote the box initially chosen, and \( O \) denote the box opened by Monty Hall. The coordinates of \( \omega = (\omega_1, \omega_2, \omega_3) \in \Omega \) correspond to \( P, S, \) and \( O \), in that order.

Now, we set our basic belief assignment, as follows:

\[
m(|x| \times |y| \times X \setminus |y|) = \frac{1}{3} p(x), \quad (43)
\]

for all \( x, y \in X \). Indeed, the position of the prize and the choice of the candidate are independent, and all we know at this point is that Monty Hall will open a box that was not chosen by the candidate.

Next, we want to condition on \( H = \{O \neq P\} \), the event that Monty Hall opens an empty box. What is the conditioned belief in \( A = \{P = S\} \)?

First, we consider contingent conditioning. For this, we need to compute \( \text{Bel}(A \cap H) \) and \( \text{Pl}(H \cap A') \). This is not difficult. The basic belief assignment in (43) only contributes to the belief in \( A \cap H \) when \( x = y \), and this leads to \( \text{Bel}(A \cap H) = 1/3 \). On the other hand, the expression does not contribute to \( A' \cap H \) precisely when \( x = y \), giving \( \text{Pl}(H \cap A') = 2/3 \). From this, we see that

\[
\text{Bel}_H(A) = \frac{1/3}{1/3 + 2/3} = 1/3,
\]

the same as the unconditional belief. On the other hand, writing \( S' = X \setminus \{S, O\} \) and \( A' = \{P = S'\} \), it is easy to compute that \( \text{Bel}_H(A') = 0 \). This becomes understandable when we realize the contingent character of the conditioning event \( H \). Indeed, it is not ruled out that, whenever \( S \neq P \), we have that \( O = P \), and thus, \( S' \neq P \).

This all changes dramatically when we condition on \( H \) in the necessary sense. Indeed, in that case we need to compute \( \text{Bel}(\{S = P\} \cup \{O = P\}) \) and \( \text{Bel}(\{S' = P\} \cup \{O = P\}) \). Given the basic belief assignment in (43), we see that only when \( x = y \), the set contributes to the belief in the former. Hence,

\[
\text{Bel}_{H, \text{ nec}}(S = P) = \frac{1}{3}.
\]

On the other hand, all sets with \( x \neq y \) contribute to the belief in \( \{S' = P\} \cup \{O = P\} \). To see this, take for instance \( x = 1 \) and \( y = 2 \) in (43). When \( O \) is either 1 or 3. If \( O = 1 \), then it contributes because \( P = 1 \). If \( O = 3 \), then \( S' = 1 \), which is equal to \( P \) and the set contributes as well.

We conclude that

\[
\text{Bel}_{H, \text{ nec}}(S' = P) = \frac{2}{3},
\]

and these numbers are in full agreement with the classical analysis of the Monty Hall problem.
5 | DS CONDITIONING

The theory of belief functions is detailed and subtle enough as to allow for various ways of conditioning. Which conditioning (or updating) rules are appropriate in a given context may depend on the interpretation one has in mind. In our interpretation, a belief function refers to a (in principle) repeatable experiment about which an agent has some, but typically not all, information. When the agent does receive information, then it is a reasonable question to know whether or not the information is about the whole experiment itself, or only about a particular outcome of one given experiment. It is very different to learn that the outcome will always be a black ball because that is how the experiment is set up, than to learn that this time the color of the ball happens to be black.

Necessary conditioning leads to adapting the basic belief assignment: The new information reveals that the experiment is actually set up differently from what the agent knew or thought before, and the new basic belief assignment accounts for precisely that. There is some transfer of probability mass taking the new information into account.

Contingent conditioning, on the other hand, applies when the agent receives information about an outcome of the experiment. The agent accounts for this by investigating what this new information means in terms of relative frequencies. As such, the new conditional beliefs are obtained by rescaling, analogous to what happens when we perform conditioning in classical probability theory. Indeed, in classical probability theory, \( P(A|B) \) is the frequency of \( A \) in the sequence of outcomes where \( B \) occurs so that outcomes that do not give \( B \) are discarded. This is what happens in contingent conditioning as well.

However, there is one major notion of conditioning in the theory of belief functions that we have not introduced so far, namely, the original DS way of conditioning introduced in Shafer (1976). We now discuss this notion of conditioning and place it into our context and interpretation.

Consider a basic belief assignment \( m \) and corresponding belief function \( \text{Bel} \). Suppose that we want to condition on the event \( H \). In the DS way of conditioning, one proceeds by assigning the basic belief assignment of a set \( A \) to the set \( A \cap H \). After this, normalization takes place by dividing through the original belief of \( H^c \) so that the sum of the new basic belief assignments is 1 again. It is not difficult to check that this leads to the belief function \( \text{Bel}_{H,\text{DS}} \) given by

\[
\text{Bel}_{H,\text{DS}}(A) = \frac{\text{Bel}(A \cup H^c) - \text{Bel}(H^c)}{1 - \text{Bel}(H^c)}
\]

with corresponding plausibility

\[
\text{Pl}_{H,\text{DS}}(A) = \frac{\text{Pl}(A \cap H)}{\text{Pl}(H)}.
\]

When we compare DS conditioning to our notions of contingent and necessary conditioning, we see that DS conditioning combines ingredients from both necessary and contingent conditioning. The transfer of basic belief assignment from \( A \) to \( A \cap H \) is a “necessary component” in the sense that the basic belief assignments are different than before. However, the subsequent rescaling is a “contingent component” in the same vein as classical and contingent conditioning. In fact, it is precisely contingent conditioning after the preceding transfer of mass in the necessary component. As such, DS conditioning is, from our point of view, a hybrid form of conditioning that is difficult to interpret in our context because it is neither necessary nor contingent, sharing characteristics with both of them.
To reinforce this point, DS conditioning can sometimes be shown to be a concatenation of necessary and contingent conditioning, as follows.

**Theorem 4.** If \( \text{Pl}(N) = 1 \) and \( \text{Bel}(H) = \text{Pl}(H) > 0 \), then

\[
\text{Bel}_{H \cap N, DS} = (\text{Bel}_{N, \text{neq}})_H = (\text{Bel}_H)_{N, \text{neq}}.
\]

**Proof.** Suppose \( \text{Pl}(N) = 1 \) and \( \text{Bel}(H) = \text{Pl}(H) \). We prove the identity for the plausibility. Because \( \text{Bel}(H) = \text{Pl}(H) \), we have

\[
\text{Pl}(A \cap H) + \text{Bel}(A^c \cap H) = \text{Pl}(H).
\]

Hence,

\[
\text{Pl}_H(A) = \frac{\text{Pl}(A \cap H)}{\text{Pl}(H)}.
\]

Applying necessary conditioning on \( N \), we find

\[
(\text{Pl}_H)_{N, \text{neq}}(A) = \frac{\text{Pl}(A \cap H \cap N)}{\text{Pl}(H)} = \text{Pl}_{N \cap H, DS}(A).
\]

If we apply the necessary conditioning on \( N \) first, we have

\[
\text{Pl}_{N, \text{neq}}(A) = \text{Pl}(A \cap N).
\]

We now observe that

\[
\text{Bel}_{N, \text{neq}}(H) = \text{Bel}(H \cup N^c)
= \text{Bel}(H)
= \text{Pl}(H)
= \text{Pl}(H \cap N)
= \text{Pl}_{N, \text{neq}}(H).
\]

Therefore, contingent conditioning on \( H \) gives

\[
(\text{Pl}_{N, \text{neq}})_H(A) = \frac{\text{Pl}_{N, \text{neq}}(A \cap H)}{\text{Pl}_{N, \text{neq}}(H)}
= \frac{\text{Pl}(A \cap H \cap N)}{\text{Pl}(H)}
= \text{Pl}_{N \cap H, DS}(A).
\]

\( \square \)

### 6 DISCUSSION AND CONCLUSIONS

In this paper, we made a distinction between what we called necessary and contingent conditioning. Compared to the focusing and updating of Dubois and Prade (1992), our caesura is slightly different. Focusing, in their setup, is frequency based, as is our contingent conditioning. Because they implicitly use the fact that the outcome could have been different, it is perhaps no surprise that they arrive at the same formula as our contingent conditioning. To be precise, they consider the same collection of consistent probability distributions as the one in Fagin and Halpern (1989), which we discussed before, that is, the collection of all probability distributions \( P \) for which \( \text{Bel}(A) \leq P(A) \leq \text{Pl}(A) \) for all \( A \subseteq \Omega \). They explain that their focusing form of conditional belief
in $A$ given $H$ can be seen as the infimum op $P(A|H)$, where the infimum runs over all consistent $P$. This is in line with the arguments in Fagin and Halpern (1989).

The updating interpretation in Dubois and Prade (1992) seems to refer to an experiment that takes place only once; the example they give is that a friend happens to give some information about the outcome of an experiment to the agent in question. If it is the case that, at least in principle, the situation is repeatable, then a frequency based interpretation still makes sense though. For instance, the Monty Hall Example 3 might be of this type. The friend (or Monty Hall) may give information either about this particular outcome or about the way the experiment is set up. In the first case, contingent conditioning still seems meaningful, and in the second case, necessary conditioning is appropriate. Hence, our caesura is different form theirs. We ask the question whether or not the information is necessarily true or contingently true, and this question seems to be relevant in all cases mentioned in Dubois and Prade (1992).

For their updating type of conditioning, Dubois and Prade arrive at the original DS form of conditioning. This form of conditioning can be derived from a special case of Dempster's rule of combination and essentially boils down to assign the mass $m(S)$ that was originally assigned to the set $S$ to the set $S \cap H$ when conditioning on $H$, followed by proper normalization.

We have explained why we think that this form of conditioning is difficult to interpret in a frequentistic framework as ours, and with our interpretation of belief functions. Dubois and Prade seem to realize this when they remark (p. 303): “Hence, as an updating rule in the frequentist framework, (5) makes sense only if [...] the piece of evidence $B$ does not contradict the set-valued statistics on $\Omega$.” (Here, equation (5) is the DS form of conditioning.) This remark is essentially the same as our earlier remark that in the definition of necessary conditioning on $H$, we need to assume that $P(\Omega) = 1$.

Nevertheless, DS conditioning pops up as a certain mixture of necessary and contingent conditioning. Given the fact that the very definition of DS conditioning is somewhat hybrid (as we have argued), this may not be a big surprise perhaps. This form of conditioning is interpretable in special circumstances, but not in all.

In this paper, we have not used Dempster’s combination rule in our derivation or motivation. Instead, we have only used frequency-based ideas, which appears to be reasonable in our interpretation of repeatable experiments about which an agent has limited information. We see the fact that our derivations do not depend on Dempster’s rule as an advantage. The rule is somewhat controversial, and we thought it would be best not to immediately lose those readers that reject it. If rules of conditioning can be derived from first principles (as we do), then we prefer this over using a rule that not all researchers endorse.

Finally, our new caesura of contingent versus necessary conditioning seems to be applicable in many situations and seems to lead to understandable answers. It relates to current notions of conditioning in the literature and approaches the concept from a new direction that we hope is found illuminating.

**ACKNOWLEDGMENT**

We thank an anonymous referee for several excellent suggestions that greatly improved the paper.

**ORCID**

*Ronald Meester* [https://orcid.org/0000-0001-7125-7864]
REFERENCES

Cohen, L. J. (1977). The probable and the provable. Oxford, England: Clarendon Press.

Cuzzolin, F. (2016). The geometry of uncertainty. Berlin, Germany: Springer.

Dietz, R. (2010). On generalizing Kolmogorov. Notre Dame Journal of Formal Logic, 51(3), 323–335.

Dubois, D., & Prade, H. (1992). Evidence, knowledge, and belief functions. International Journal of Approximate Reasoning, 6(3), 295–319.

Fagin, R., & Halpern, J. Y. (1989). Uncertainty, belief and probability. In Proceedings of the International Joint Conference on AI, Detroit, MI, 1, pp. 1161–1167.

Haenni, R. (2009). Non-additive degrees of belief (pp. 121–159). Dordrecht, The Netherlands: Springer.

Jaffray, J.-Y. (1990). Bayesian updating and belief functions. Proceedings of the 3rd International Conference IMPU’90, Paris, 449–451.

Kerkvliet, T., & Meester, R. (2016). Assessing forensic evidence by computing belief functions. Law, Probability and Risk, 15(2), 127–153.

Marinacci, M. (1999). Limit laws for non-additive probabilities and their frequentist interpretation. Journal of Economic Theory, 84, 145–195.

Pearl, J. (1990). Reasoning with belief functions: An analysis of compatibility. International Journal of Approximate Reasoning, 4, 363–389.

Royall, R. (1997). Statistical evidence: A likelihood paradigm. Boca Raton, FL: Chapman & Hall.

Shafer, G. (1976). A mathematical theory of evidence. Princeton, NJ: Princeton University Press.

Shafer, G. (1981). Constructive probability. Synthese, 48(1), 1–60.

Shafer, G. (2008). Non-additive probabilities in the work of Bernoulli and Lambert. In R. Yager & L. Liu (Eds.), Classic works of the Dempster-Shafer theory of belief functions (pp. 117–181). Berlin, Germany: Springer.

Walley, P. (1991). Statistical reasoning with imprecise probabilities. Boca Raton, FL: Chapman & Hall.

Walley, P., & Fine, T. L. (1982). Towards a frequentist theory of upper and lower probability. The Annals of Statistics, 10, 741–761.

How to cite this article: Meester R, Kerkvliet T. A new look at conditional probability with belief functions. Statistica Neerlandica. 2019;73:274–291. https://doi.org/10.1111/stan.12169