Bayesian Inference under Ambiguity: Conditional Prior Belief Functions

Giulianella Coletti GIULIANELLA.COLETTI@UNIPG.IT
Dip. Matematica e Informatica, Università degli Studi di Perugia (Italy)

Davide Petturiti DAVIDE.PETTURITI@UNIPG.IT
Dip. Economia, Università degli Studi di Perugia (Italy)

Barbara Vantaggi BARBARA.VANTAGGI@SBAI.UNIROMA1.IT
Dip. S.B.A.I., "La Sapienza" Università di Roma (Italy)

Abstract

Bayesian inference under imprecise prior information is studied: the starting point is a precise strategy σ and a full B-conditional prior belief function Bel_B, conveying ambiguity in probabilistic prior information. In finite spaces, we give a closed form expression for the lower envelope $\underline{P}$ of the class of full conditional probabilities dominating $\{Bel_B, \sigma\}$ and, in particular, for the related "posterior probabilities". The assessment $\{Bel_B, \sigma\}$ is a coherent lower conditional probability in the sense of Williams and the characterized lower envelope $\underline{P}$ coincides with its natural extension.

Keywords: conditional belief function; Bayesian conditioning rule; inference; ambiguity.

1. Introduction

Bayesian inference is known to naturally fit into de Finetti's theory of coherent (finitely additive) conditional probabilities, where a coherent assessment can be always extended, generally not in a unique way, to any superset of conditional events (de Finetti, 1975; Williams, 1975).

In some application domains (e.g., decision theory, economics, game theory and forensic analysis, to cite some) the prior knowledge could be only partially specified or, even worse, it could refer to a different space of hypotheses. In these circumstances, instead of considering a single prior distribution, one is forced to take into account a set of priors (see, e.g., (Dempster, 1967; DeRoberts and Hartigan, 1981; Gilboa and Schmeidler, 1989)).

For instance, suppose that a pension system, based on the social security contributions Λ, is modified by a legal reform so that the new pension scheme takes into account the contribution's years Θ. In order to use the previous information, we need to extract a new prior, starting from the prior distribution P of Λ by taking into account the logical relations between Λ and Θ. There could be possibly infinite probability distributions of (Λ, Θ) compatible with P, determining a lower envelope for the distribution of Θ. In particular, if the initial prior information P is a full conditional probability (Dubins, 1975), then for Θ we obtain a *full B-conditional belief function* (Coletti et al., 2016b), i.e., a conditional totally monotone uncertainty measure. Now, considering the profession X and a statistical model connecting X and Θ, the goal could be to draw inferences on Θ belonging to a set of values A (e.g., social pension) under particular values of X (e.g., a person is a clerk).

Motivated by the previous discussion, the main aim of this paper is to prove a generalized version of Bayes' theorem, working with an ambiguous conditional prior information in the form of a full B-conditional belief function Bel_B and a precise statistical model λ, the latter uniquely determining a strategy σ (Dubins, 1975). A prior in the form of a full B-conditional belief function is not so uncommon. For instance, in Example 2, starting from an automatic system $\mathbf{S}$ which evolves according to a Markov chain, we show how to generate a full B-conditional prior belief function on the algebra spanned by the states of another unobservable automatic system $\mathbf{T}$, taking into account the logical constraints among the states of $\mathbf{S}$ and those of $\mathbf{T}$.

Focusing on finite spaces, we provide a characterization of the lower envelope $\underline{P}$ of the class of full conditional probabilities dominating $\{Bel_B, \sigma\}$. The assessment $\{Bel_B, \sigma\}$ is a Williams-coherent lower conditional probability and $\underline{P}$ turns out to be its natural extension (Williams, 1975).

Our results are connected with those proved in (Walley, 1981, 1991; Wasserman, 1990a,b; Wasserman and Kadane, 1990): we generalize them in a finite context, since no assumption of positivity for the (lower or upper) probability of the conditioning events is required.

2. Preliminaries

Let $\mathcal{A}$ be a Boolean algebra of *events* E's, and denote with $(\cdot)^c$, $\vee$ and $\wedge$ the usual Boolean operations of negation, disjunction and conjunction, respectively, and with $\subseteq$ the partial order of implication. The *sure event* Ω and the *impossible event* $\emptyset$ coincide, respectively, with the top and bottom elements of $\mathcal{A}$. If $\mathcal{A}$ is finite, we denote with $\mathcal{C}_\mathcal{A}$ the subset of its *atoms* which form the finer partition of Ω contained in $\mathcal{A}$. Denote $\mathcal{A}^0 = \mathcal{A} \setminus \{\emptyset\}$, $\mathbb{N}$ is the set of natural numbers, I stands for an arbitrary index set and $\langle \{E_i\}_{i\in I}\rangle$ indicates the Boolean algebra generated by the set of events $\{E_i\}_{i\in I}$.

A *conditional event* $E|H$ is an ordered pair of events (E, H) with $H \neq \emptyset$. In particular, any event E can be identified with the conditional event $E|\Omega$. An arbitrary set of conditional events $\mathcal{G} = \{E_i|H_i\}_{i\in I}$ can always be embedded into a minimal set $\mathcal{A} \times \mathcal{A}^0$, where $\mathcal{A} = \langle \{E_i, H_i\}_{i\in I}\rangle$.

Recall the definition of *coherent conditional probability* essentially due to (de Finetti, 1975; Holzer, 1984; Regazzini, 1985; Williams, 1975).

Definition 1 *Let $\mathcal{G} = \{E_i|H_i\}_{i\in I}$ be a set of conditional events. A function $P : \mathcal{G} \to [0, 1]$ is a* **coherent conditional probability** *if and only if, for every $n \in \mathbb{N}$, every $E_{i_1}|H_{i_1}, \dots, E_{i_n}|H_{i_n} \in \mathcal{G}$ and every real numbers $s_1, \dots, s_n$, denoting $\mathcal{B} = \langle \{E_{i_j}, H_{i_j}\}_{j=1,\dots,n}\rangle$ with set of atoms $\mathcal{C}_\mathcal{B} = \{C_1, \dots, C_m\}$, the random gain defined on $\mathcal{C}_\mathcal{B}$ as $G = \sum_{j=1}^n s_j(\mathbf{1}_{E_{i_j}} - P(E_{i_j}|H_{i_j}))\mathbf{1}_{H_{i_j}}$ satisfies*

$$\min_{C_r \subseteq H_0^0} G(C_r) \leq 0 \leq \max_{C_r \subseteq H_0^0} G(C_r),$$

where $H_0^0 = \bigvee_{j=1}^n H_{i_j}$ and, for every $E \in \mathcal{B}$, $\mathbf{1}_E$ is its indicator defined on $\mathcal{C}_\mathcal{B}$ as $\mathbf{1}_E(C_r) = 1$ if $C_r \subseteq E$ and 0 otherwise.

In particular, if $\mathcal{G} = \mathcal{A} \times \mathcal{A}^0$ where $\mathcal{A}$ is a Boolean algebra, then $P(\cdot|\cdot)$ is a coherent conditional probability if and only if it satisfies the following conditions:

(C1) $P(E|H) = P(E \wedge H|H)$, for every $E \in \mathcal{A}$ and $H \in \mathcal{A}^0$;

(C2) $P(\cdot|H)$ is a finitely additive probability on $\mathcal{A}$, for every $H \in \mathcal{A}^0$;

(C3) $P(E \wedge F|H) = P(E|H) \cdot P(F|E \wedge H)$, for every $H, E \wedge H \in \mathcal{A}^0$ and $E, F \in \mathcal{A}$.

In this case $P(\cdot|\cdot)$ is simply said a *full conditional probability on* $\mathcal{A}$ according to (Dubins, 1975).

If $\mathcal{G} = \{E_i|H_i\}_{i\in I}$ is an arbitrary set, the coherence condition is equivalent to the existence of a full conditional probability on the $\mathcal{A} = \langle \{E_i, H_i\}_{i\in I}\rangle$ extending the given assessment. This is a consequence of the conditional version of the *fundamental theorem for probabilities* (de Finetti, 1975; Regazzini, 1985; Williams, 1975): every coherent conditional probability P on an arbitrary $\mathcal{G}$ can be extended, generally not in a unique way, to every superset of conditional events $\mathcal{G}'$.

If $\mathcal{A}$ is a finite Boolean algebra, every full conditional probability $P(\cdot|\cdot)$ on $\mathcal{A}$ is in bijection with a unique linearly ordered class $\{P_0, \ldots, P_k\}$ of probability measures on $\mathcal{A}$ whose supports form a partition of Ω (Krauss, 1968). The class $\{P_0, \ldots, P_k\}$ is called *complete agreeing class* and represents the full conditional probability $P(\cdot|\cdot)$ in the sense that, for every $F|K \in \mathcal{A} \times \mathcal{A}^0$, there is a minimum index $\alpha \in \{0, \ldots, k\}$ such that $P_\alpha(K) > 0$ and $P(F|K) = \frac{P_\alpha(F \wedge K)}{P_\alpha(K)}$.

If $\mathcal{G}$ is an arbitrary set, we can have more complete agreeing classes, each of them obtained by solving a suitable sequence of linear systems (Coletti and Scozzafava, 2002).

The set $\mathcal{P} = \{\tilde{P}(\cdot|\cdot)\}$ of all coherent extensions of P to a superset $\mathcal{G}'$ is a compact subset of the space $[0,1]^{\mathcal{G}'}$ endowed with the product topology and the projection set on each element of $\mathcal{G}'$ is a (possibly degenerate) closed interval. The pointwise envelopes

$$\underline{P} = \min \mathcal{P} \quad \text{and} \quad \overline{P} = \max \mathcal{P},$$

are known as *coherent lower* and *upper conditional probabilities* (Coletti and Scozzafava, 2002), where coherence here is intended in the sense of (Williams, 1975) (namely, *Williams-coherence*). The envelopes $\underline{P}$ and $\overline{P}$ satisfy the *duality* property, i.e., $\overline{P}(E|H) = 1 - \underline{P}(E^c|H)$, for every $E|H, E^c|H \in \mathcal{G}'$, so in the following we mainly deal with $\underline{P}$.

In general, Williams-coherent lower conditional probabilities can be introduced without starting from a precise coherent conditional probability (Williams, 1975):

Definition 2 *A function $\underline{P}(\cdot|\cdot)$ on a set of conditional events $\mathcal{G} = \{E_i|H_i\}_{i \in I}$ is a* **Williams-coherent lower conditional probability** *if there is a class of $\mathcal{P} = \{\tilde{P}(\cdot|\cdot)\}$ of coherent conditional probabilities on $\mathcal{G}$ such that $\underline{P} = \inf \mathcal{P}$.*

The extendibility of every coherent conditional probability implies the extendibility, generally not in a unique way, of every Williams-coherent lower conditional probability: the pointwise minimal of such extension is referred to as *natural extension* (Williams, 1975). For checking that an assessment is Williams-coherent in a finite setting and to find its natural extension see (Capotorti et al., 2003; Coletti and Scozzafava, 2002).

3. Full B-conditional belief and plausibility functions

A *belief function* Bel (Dempster, 1967; Shafer, 1976) on a finite Boolean algebra $\mathcal{A}$ with set of atoms $\mathcal{C}_\mathcal{A}$ is a function such that $Bel(\emptyset) = 0$, $Bel(\Omega) = 1$ and satisfying the n*-monotonicity* property for every $n \geq 2$, i.e., for every $E_1, \ldots, E_n \in \mathcal{A}$,

$$Bel\left(\bigvee_{i=1}^{n} E_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel\left(\bigwedge_{i \in I} E_i\right).$$

The associated dual function Pl, defined as $Pl(E) = 1 - Bel(E^c)$, for every $E \in \mathcal{A}$, is said *plausibility function*. Both Bel and Pl are particular *(normalized) capacities* (Choquet, 1953), i.e., they are monotone with respect to the $\subseteq$ relation. A belief function Bel (and so its dual Pl) on a finite Boolean algebra is completely characterized by its Möbius inversion $m : \mathcal{A} \to [0,1]$, also known as *basic probability assignment* (Shafer, 1976), defined, for every $E \in \mathcal{A}$, as

$$m(E) = \sum_{B \subseteq E} (-1)^{|\mathcal{C}_{E \setminus B}|} Bel(B),$$

where $\mathcal{C}_{E\setminus B} = \{D_r \in \mathcal{C}_{\mathcal{A}} : D_r \subseteq E \wedge B^c\}$. In particular, m satisfies $m(\emptyset) = 0$ and $\sum_{E\in\mathcal{A}} m(E) = 1$, and, for every $E \in \mathcal{A}$, it holds

$$Bel(E) = \sum_{B \subseteq E} m(E) \quad \text{and} \quad Pl(E) = \sum_{B \wedge E \neq \emptyset} m(E).$$

Denote with $\mathbf{F}_{Bel}$ the set of *focal elements* of Bel, where an event A in $\mathcal{A}$ is a focal element of Bel whenever $m(A) > 0$.

Given a finite partition $\mathcal{L} = \{H_1, \ldots, H_n\}$ of Ω, a capacity φ on $\mathcal{A}_{\mathcal{L}} = \langle \mathcal{L} \rangle$ and a function $X : \mathcal{L} \to \mathbb{R}$, the *Choquet integral* of X with respect to φ is defined as

$$\oint X(H_i)\varphi(\mathrm{d}H_i) = \sum_{i=1}^{n} \left[X(H_{\rho(i)}) - X(H_{\rho(i-1)})\right] \varphi\left(H_{\rho(i)} \vee \ldots \vee H_{\rho(n)}\right),$$

where ρ is a permutation of $\{1, \ldots, n\}$ such that $X(H_{\rho(1)}) \leq \ldots \leq X(H_{\rho(n)})$ and $X(H_{\rho(0)}) := 0$. We write $\mathrm{d}H_i$ since we are integrating a function defined on the partition $\mathcal{L} = \{H_1, \ldots, H_n\}$ with respect to a capacity defined on $\mathcal{A}_{\mathcal{L}} = \langle \mathcal{L} \rangle$.

We recall the definitions of C-class and full B-conditional belief and plausibility functions given in (Coletti et al., 2016b).

Definition 3 *Let $\mathcal{A}$ be a finite Boolean algebra. A linearly ordered class $\{Bel_0, \ldots, Bel_k\}$ of belief functions on $\mathcal{A}$ with sets of focal elements $\{\mathbf{F}_{Bel_0}, \ldots, \mathbf{F}_{Bel_k}\}$ is said a* **covering class**, *or* **C-class** *for short, if it satisfies the following covering condition*

$$\bigvee_{E \in \bigcup_{\alpha=0}^{k} \mathbf{F}_{Bel_\alpha}} E = \Omega. \tag{1}$$

By duality, a linearly ordered class $\{Pl_0, \ldots, Pl_k\}$ of plausibility functions on $\mathcal{A}$ is said *C-class* if the corresponding class of dual belief functions $\{Bel_0, \ldots, Bel_k\}$ is. By means of a C-class of belief functions we define the concept of full B-conditional belief function, where B stands for Bayesian.

Definition 4 *Let $\mathcal{A}$ be a finite Boolean algebra. A function $Bel_B : \mathcal{A} \times \mathcal{A}^0 \to [0,1]$ is a* **full B-conditional belief function** *on $\mathcal{A}$ if there exists a C-class $\{Bel_0, \ldots, Bel_k\}$ of belief functions on $\mathcal{A}$ whose dual plausibility functions are $\{Pl_0, \ldots, Pl_k\}$, such that, for every $E|H \in \mathcal{A} \times \mathcal{A}^0$, if $E \wedge H = H$ then $Bel_B(E|H) = 1$, while if $E \wedge H \neq H$*

$$Bel_B(E|H) = \frac{Bel_{\alpha_{E,H}}(E \wedge H)}{Bel_{\alpha_{E,H}}(E \wedge H) + Pl_{\alpha_{E,H}}(E^c \wedge H)}, \tag{2}$$

where $\alpha_{E,H} = \min\{\alpha \in \{0, \ldots, k\} \,:\, Bel_\alpha(E \wedge H) + Pl_\alpha(E^c \wedge H) > 0\}$.

The previous definition introduces a full B-conditional belief function through a *generalized Bayesian conditioning rule* corresponding to the one originally given in (Walley, 1981) for 2-monotone capacities. The Bayesian conditioning rule has been discussed for belief functions in (Dempster, 1967; Dubois and Denœux, 2012; Fagin and Halpern, 1991; Jaffray, 1992).

The difference with the previous ones is that the rule given in Definition 4 covers also the case in which $Bel(E \wedge H|\Omega) + Pl(E^c \wedge H|\Omega) = Bel_0(E \wedge H) + Pl_0(E^c \wedge H) = 0$, since it relies not on a single belief function but on a linearly ordered class of belief functions.

For $H \in \mathcal{A}^0$, the dual conditional function Pl_B of a full B-conditional belief function Bel_B on $\mathcal{A}$ is defined, for every $E \in \mathcal{A}$, as

$$Pl_B(E|H) = 1 - Bel_B(E^c|H),$$

and is called *full B-conditional plausibility function*. By duality we immediately have $Pl_B(E|H) = 0$ when $E \wedge H = \emptyset$, while if $E \wedge H \neq \emptyset$ it holds

$$Pl_B(E|H) = 1 - Bel_B(E^c|H) = \frac{Pl_{\alpha_{E^c,H}}(E \wedge H)}{Pl_{\alpha_{E^c,H}}(E \wedge H) + Bel_{\alpha_{E^c,H}}(E^c \wedge H)}. \tag{3}$$

Notice that a full conditional probability P on $\mathcal{A}$ is both a full B-conditional belief function and a full B-conditional plausibility function.

In this paper conditional belief functions are always intended in the sense of Definition 4: notice that full conditional probabilities reveal to be both full B-conditional belief and full B-conditional plausibility functions.

In (Coletti et al., 2016b) it is proved that the conditional measures Bel_B and Pl_B determine the non-empty compact set

$$\mathcal{P}_B = \{\tilde{\pi} \ : \ \tilde{\pi} \text{ is a full conditional probability on } \mathcal{A},\, Bel_B \leq \tilde{\pi} \leq Pl_B\}, \tag{4}$$

for which it holds $Bel_B = \min \mathcal{P}_B$ and $Pl_B = \max \mathcal{P}_B$. In the same paper we prove that, for every full B-conditional belief function Bel_B on $\mathcal{A}$ it is always possible to find a different finite Boolean algebra $\mathcal{B}$ and a full conditional probability P on $\mathcal{B}$ such that $\mathcal{P}_B$ can be recovered as the set of coherent extensions of P to $\mathcal{A} \times \mathcal{A}^0$ and, thus, Bel_B and Pl_B as the envelopes of $\mathcal{P}_B$.

In (Coletti et al., 2016a) it is also shown that if all the belief functions in a C-class reduce to necessity measures then the corresponding full B-conditional belief function is a *full B-conditional necessity measure* and its dual is a *full B-conditional possibility measure*. In particular, interpreting the conditional measures Bel_B and Pl_B as envelopes, a necessary and sufficient condition (involving the finite Boolean algebras $\mathcal{B}$ and $\mathcal{A}$ and the full conditional probability P) is given for Bel_B and Pl_B to be full B-conditional necessity and possibility measures.

Since a full B-conditional belief function Bel_B determines the non-empty compact set $\mathcal{P}_B$ of full conditional probabilities dominating it, its use in a Bayesian inferential procedure implies an ambiguous specification of a full conditional probability.

4. Bayesian inference with full B-conditional prior belief functions

Let $\mathcal{L} = \{H_1, \ldots, H_n\}$ and $\mathcal{E} = \{E_1, \ldots, E_m\}$ be two finite partitions of Ω and consider the Boolean algebras $\mathcal{A}_\mathcal{L} = \langle \mathcal{L} \rangle$, $\mathcal{A}_\mathcal{E} = \langle \mathcal{E} \rangle$, $\mathcal{A} = \langle \mathcal{A}_\mathcal{L} \cup \mathcal{A}_\mathcal{E} \rangle$. The partitions $\mathcal{L}$ and $\mathcal{E}$ play the roles of the sets of mutually exclusive and exhaustive "hypotheses" and "evidences", respectively.

In the standard Bayesian setting, a *statistical model* (see, e.g., (Torgersen, 1991)) is given on $\mathcal{A}_{\mathcal{E}} \times \mathcal{L}$, where the latter is a function $\lambda : \mathcal{A}_{\mathcal{E}} \times \mathcal{L} \to [0,1]$ such that, for every $H_i \in \mathcal{L}$:

(L1) $\lambda(B|H_i) = 0$ if $B \wedge H_i = \emptyset$ and $\lambda(B|H_i) = 1$ if $B \wedge H_i = H_i$, for every $B \in \mathcal{A}_{\mathcal{E}}$;

(L2) $\lambda(\cdot|H_i)$ is a probability on $\mathcal{A}_{\mathcal{E}}$.

Proposition 1 in (Petturiti and Vantaggi, 2017) implies that any statistical model λ on $\mathcal{A}_{\mathcal{E}} \times \mathcal{L}$ uniquely extends to a *strategy* on $\mathcal{A} \times \mathcal{L}$ (see, e.g., (Dubins, 1975)) which is a function $\sigma : \mathcal{A} \times \mathcal{L} \to [0,1]$ such that, for every $H_i \in \mathcal{L}$:

(S1) $\sigma(H_i|H_i) = 1$;

(S2) $\sigma(\cdot|H_i)$ is a probability on $\mathcal{A}$.

By Theorem 5 in (Dubins, 1975), for every full conditional prior probability π on $\mathcal{A}_{\mathcal{L}}$, the assessment $\{\pi, \sigma\}$ (and, in particular, $\{\pi, \lambda\}$) is a coherent conditional probability, therefore it can be extended, generally not in a unique way, to a full conditional probability on $\mathcal{A}$. This implies that, given a full B-conditional prior belief function Bel_B on $\mathcal{A}$, the assessment $\{Bel_B, \sigma\}$ (and, in particular, $\{Bel_B, \lambda\}$) is a Williams-coherent lower conditional probability.

Remark 5 *The assessment $\{Bel_B, \sigma\}$ determines a Williams-coherent lower conditional probability $\underline{P}$ on the set of conditional events $\mathcal{G} = (\mathcal{A}_{\mathcal{L}} \times \mathcal{A}^0_{\mathcal{L}}) \cup (\mathcal{A} \times \mathcal{L})$ such that $\underline{P}_{|\mathcal{A}_{\mathcal{L}} \times \mathcal{A}^0_{\mathcal{L}}} = Bel_B$ and $\underline{P}_{\mathcal{A} \times \mathcal{L}} = \sigma$ so, with a little abuse of terminology, $\{Bel_B, \sigma\}$ is directly referred to be a Williams-coherent lower conditional probability.*

Remark 6 *Recall that, in view of the finite setting adopted in this paper, the notion of conditioning for lower probabilities due to Williams coincides with that due to (Walley, 1991) since in this case the conglomerability condition is automatically satisfied.*

Let Bel_B be a full B-conditional belief function on $\mathcal{A}_{\mathcal{L}}$ and σ a strategy on $\mathcal{A} \times \mathcal{L}$ and denote with $\mathcal{P}_B$ the set of full conditional probabilities on $\mathcal{A}_{\mathcal{L}}$ dominating Bel_B. Consider

$$\mathcal{P} = \{\tilde{P} \ : \ \tilde{P} \text{ is a full conditional probability on } \mathcal{A} \text{ extending } \{\tilde{\pi}, \sigma\}, \tilde{\pi} \in \mathcal{P}_B\},$$

which is a non-empty compact subset of $[0,1]^{\mathcal{A} \times \mathcal{A}^0}$ endowed with the product topology, whose envelopes are $\underline{P} = \min \mathcal{P}$ and $\overline{P} = \max \mathcal{P}$. The lower envelope $\underline{P}(\cdot|\cdot)$ turns out to be the natural extension of the Williams-coherent lower conditional probability $\{Bel_B, \sigma\}$.

The following theorem provides a characterization of the lower envelope $\underline{P}(\cdot|\cdot)$, relying on the functions defined, for every $F, K \in \mathcal{A}$ and $A \in \mathcal{A}^0_{\mathcal{L}}$ such that $K \subseteq A$, as

$$L(F, K; A) = \min_{\tilde{\pi} \in \mathcal{P}_B} \left\{ \sum_{i=1}^n \sigma(FK|H_i)\tilde{\pi}(H_i|A) : \sum_{i=1}^n \sigma(F^cK|H_i)\tilde{\pi}(H_i|A) = \overline{P}(F^cK|A) \right\},$$

$$U(F, K; A) = \max_{\tilde{\pi} \in \mathcal{P}_B} \left\{ \sum_{i=1}^n \sigma(FK|H_i)\tilde{\pi}(H_i|A) : \sum_{i=1}^n \sigma(F^cK|H_i)\tilde{\pi}(H_i|A) = \underline{P}(F^cK|A) \right\},$$

where we write FK and F^cK in place of $F \wedge K$ and $F^c \wedge K$ to save space.

Theorem 7 *The lower envelope $\underline{P}(\cdot|\cdot)$ is such that, for every $F|K \in \mathcal{A} \times \mathcal{A}^0$, if $F \wedge K = K$, then $\underline{P}(F|K) = 1$, otherwise:*

(i) if $K \in \mathcal{A}^0_{\mathcal{L}}$, then
$$\underline{P}(F|K) = \oint \sigma(F|H_i) Bel_B(\mathrm{d}H_i|K);$$

(ii) if $K \in \mathcal{A}^0 \setminus \mathcal{A}^0_{\mathcal{L}}$, then if there exists $A \in \mathcal{A}^0_{\mathcal{L}}$ such that $K \subseteq A$ and $\underline{P}(K|A) > 0$ we have that
$$\underline{P}(F|K) = \min\left\{\frac{\underline{P}(F \wedge K|A)}{\underline{P}(F \wedge K|A) + U(F^c, K; A)}, \frac{L(F, K; A)}{L(F, K; A) + \overline{P}(F^c \wedge K|A)}\right\},$$

otherwise $\underline{P}(F|K) = 0$.

Proof The statement is trivial if $F \wedge K = K$ since in this case $\tilde{P}(F|K) = 1$ for every $\tilde{P} \in \mathcal{P}$, thus suppose $F \wedge K \neq K$. Condition *(i)* follows since, if $K \in \mathcal{A}^0_{\mathcal{L}}$ then

$$\begin{aligned}
\underline{P}(F|K) &= \min_{\tilde{P}\in\mathcal{P}} \tilde{P}(F|K) = \min_{\tilde{\pi}\in\mathcal{P}_B} \sum_{i=1}^{n} \sigma(F|H_i)\tilde{\pi}(H_i|K) \\
&= \min_{\tilde{\pi}\in\mathcal{C}_{Bel_B(\cdot|K)}} \sum_{i=1}^{n} \sigma(F|H_i)\tilde{\pi}(H_i|K) = \oint \sigma(F|H_i) Bel_B(\mathrm{d}H_i|K),
\end{aligned}$$

where $\mathcal{C}_{Bel_B(\cdot|K)} = \{\tilde{\pi}(\cdot|K)\}$ is the core of probability measures on $\mathcal{A}_{\mathcal{L}}$ induced by $Bel_B(\cdot|K)$ and the last equality follows by Proposition 3 in (Schmeidler, 1986). To prove condition *(ii)* let us consider $K \in \mathcal{A}^0 \setminus \mathcal{A}^0_{\mathcal{L}}$. If there exists $A \in \mathcal{A}^0_{\mathcal{L}}$ such that $K \subseteq A$ and $\underline{P}(K|A) > 0$ we have that $\tilde{P}(K|A) > 0$ for every $\tilde{P} \in \mathcal{P}$, thus $\underline{P}(F|K) = \min_{\tilde{P}\in\mathcal{P}} \frac{\tilde{P}(F\wedge K|A)}{\tilde{P}(F\wedge K|A)+\tilde{P}(F^c\wedge K|A)}$. So, the conclusion follows since $\frac{x}{x+y}$ is increasing in x and decreasing in y, and, for every $\tilde{P} \in \mathcal{P}$, $\tilde{P}(\cdot|A)$ is the convex combination of probabilities $P_1(\cdot|A)$ and $P_2(\cdot|A)$, $P_1, P_2 \in \mathcal{P}$, attaining the lower and the upper envelopes, respectively, on $F \wedge K$ (or on $F^c \wedge K$) and $\tilde{P}(F|K) \geq \min\{P_1(F|K), P_2(F|K)\}$. The remaining case, realizing when for all $A \in \mathcal{A}^0_{\mathcal{L}}$ with $K \subseteq A$ it holds $\underline{P}(K|A) = 0$, is proven in analogy to the proof of Lemma 3 in (Petturiti and Vantaggi, 2017). ■

Restricting to a finite setting, the previous theorem generalizes some results proved in (Coletti et al., 2014) in which an ambiguous unconditional prior is considered, either in the form of a belief function or a 2-monotone capacity.

A simplification of condition *(ii)* of Theorem 7 is obtained when the functions on $\mathcal{L}$, defined as $X(\cdot) = \sigma(F \wedge H|\cdot)$ and $(1 - Y(\cdot)) = (1 - \sigma(F^c \wedge H|\cdot))$, are *comonotonic* (see, e.g., (Denneberg, 1994)), i.e., for every $H_h, H_k \in \mathcal{L}$, $[X(H_h) - X(H_k)] \cdot [(1 - Y(H_h)) - (1 - Y(H_k))] \geq 0$, as shown by the following Proposition 8. In particular, this happens for all conditional events in $\mathcal{A}_{\mathcal{L}} \times \mathcal{A}^0_{\mathcal{E}}$ related to "posterior probabilities", obtaining, for a finite setting, a generalization of results in (Wasserman, 1990a).

Proposition 8 *For every $F|K \in \mathcal{A} \times \mathcal{A}^0$ such that $F \wedge K \neq K$, $K \in \mathcal{A}^0 \setminus \mathcal{A}^0_{\mathcal{L}}$ and there exists $A \in \mathcal{A}^0_{\mathcal{L}}$ such that $K \subseteq A$ and $\underline{P}(K|A) > 0$, if $X(\cdot) = \sigma(F \wedge H|\cdot)$ and $(1 - Y(\cdot)) = (1 - \sigma(F^c \wedge H|\cdot))$ are comonotonic then*
$$\underline{P}(F|K) = \frac{\underline{P}(F \wedge K|A)}{\underline{P}(F \wedge K|A) + \overline{P}(F^c \wedge K|A)}.$$

Proof For every $A \in \mathcal{A}^0_{\mathcal{L}}$, $Bel_B(\cdot|A)$ is a totally monotone capacity on $\mathcal{A}_{\mathcal{L}}$ inducing a core $\mathcal{C}_{Bel_B(\cdot|A)} = \{\tilde{\pi}(\cdot|A)\}$ of probability measures on $\mathcal{A}_{\mathcal{L}}$, moreover, the functions $X(\cdot)$ and $(1-Y(\cdot))$ are comonotonic.

By Proposition 6.26 in (Troffaes and de Cooman, 2014) there exists $\tilde{\pi}(\cdot|A) \in \mathcal{C}_{Bel_B(\cdot|A)}$ such that $\sum_{i=1}^n X(H_i)\tilde{\pi}(H_i|A) = \oint X(H_i)Bel_B(\mathrm{d}H_i|A)$ and $\sum_{i=1}^n (1-Y(H_i))\tilde{\pi}(H_i|A) = \oint(1-Y(H_i))Bel_B(\mathrm{d}H_i|A)$.

Since $\sum_{i=1}^n (1-Y(H_i))\tilde{\pi}(H_i|A) = 1-\sum_{i=1}^n Y(H_i)\tilde{\pi}(H_i|A)$ and $\oint(1-Y(H_i))Bel_B(\mathrm{d}H_i|A) = 1-\oint Y(H_i)Pl_B(\mathrm{d}H_i|A)$, it follows $\oint Y(H_i)Pl_B(\mathrm{d}H_i|A) = \sum_{i=1}^n Y(H_i)\tilde{\pi}(H_i|A)$ and this implies $\underline{P}(F \wedge K|A) = \oint X(H_i)Bel_B(\mathrm{d}H_i|A) = \sum_{i=1}^n X(H_i)\tilde{\pi}(H_i|A) = L(F,K;A)$ and $\overline{P}(F^c \wedge K|A) = \oint Y(H_i)Pl_B(\mathrm{d}H_i|A) = \sum_{i=1}^n Y(H_i)\tilde{\pi}(H_i|A) = U(F^c,K;A)$. ■

The following example shows that, though $Bel_B(\cdot|K)$ is a belief function on $\mathcal{A}_{\mathcal{L}}$, for every $K \in \mathcal{A}^0_{\mathcal{L}}$, and $\sigma(\cdot|H_i)$ is a probability measure on $\mathcal{A}$, for every $H_i \in \mathcal{L}$, the function $\underline{P}(\cdot|K)$ can fail 2-monotonicity, for some $K \in \mathcal{A}^0$.

Example 1 *Let $\mathcal{L} = \{H_1, H_2, H_3\}$ and $\mathcal{E} = \{E_1, E_2, E_3, E_4\}$ be logically independent partitions of Ω, and take $\mathcal{A}_{\mathcal{L}} = \langle\mathcal{L}\rangle$, $\mathcal{A}_{\mathcal{E}} = \langle\mathcal{E}\rangle$ and $\mathcal{A} = \langle\mathcal{A}_{\mathcal{L}} \cup \mathcal{A}_{\mathcal{E}}\rangle$. Let Bel_B be the full B-conditional belief function on $\mathcal{A}_{\mathcal{L}}$ determined by the C-class of belief functions $\{Bel_0, Bel_1\}$ displayed below*

$\mathcal{A}_{\mathcal{L}}$	$\emptyset$	H_1	H_2	H_3	$H_1 \vee H_2$	$H_1 \vee H_3$	$H_2 \vee H_3$	Ω
Bel_0	0	$\frac{1}{2}$	0	0	1	$\frac{1}{2}$	0	1
Bel_1	0	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1

where $\mathbf{F}_{Bel_0} = \{H_1, H_1 \vee H_2\}$ and $\mathbf{F}_{Bel_1} = \{H_1, H_2 \vee H_3\}$, thus condition (1) of Definition 3 is satisfied.

Let λ be the statistical model on $\mathcal{A}_{\mathcal{E}} \times \mathcal{L}$ such that

$$\lambda(E_j|H_1) = \lambda(E_j|H_3) = \frac{1}{6}, \text{ for } j = 1,2,3, \text{ and } \lambda(E_4|H_1) = \lambda(E_4|H_3) = \frac{1}{2},$$

$$\lambda(E_1|H_2) = \lambda(E_3|H_2) = \frac{1}{2}, \text{ and } \lambda(E_2|H_2) = \lambda(E_4|H_2) = 0.$$

which uniquely extends to a strategy σ on $\mathcal{A} \times \mathcal{L}$ by Proposition 1 in (Petturiti and Vantaggi, 2017). Let $K = H_2 \vee H_3$, $A = E_1 \vee E_2$ and $B = E_2 \vee E_3$. Simple computations show that $Bel_B(\cdot|K)$ is a belief function vacuous at K, so, we have

$$\begin{aligned}
\underline{P}(A \vee B|K) &= \oint \sigma(A \vee B|H_i)Bel_B(\mathrm{d}H_i|K) = \inf_{H_i \subseteq K} \sigma(A \vee B|K) = \frac{1}{2},\\
\underline{P}(A|K) &= \oint \sigma(A|H_i)Bel_B(\mathrm{d}H_i|K) = \inf_{H_i \subseteq K} \sigma(A|K) = \frac{1}{3},\\
\underline{P}(B|K) &= \oint \sigma(B|H_i)Bel_B(\mathrm{d}H_i|K) = \inf_{H_i \subseteq K} \sigma(B|K) = \frac{1}{3},\\
\underline{P}(A \wedge B|K) &= \oint \sigma(A \wedge B|H_i)Bel_B(\mathrm{d}H_i|K) = \inf_{H_i \subseteq K} \sigma(A \wedge B|K) = 0.
\end{aligned}$$

Since $\underline{P}(A \vee B|K) < \underline{P}(A|K) + \underline{P}(B|K) - \underline{P}(A \wedge B|K)$, $\underline{P}(\cdot|K)$ is not 2-monotone.

Proposition 8 is a generalization of the ϵ-contamination model presented in Example 2.3 in (Huber, 1981), where the author provides a characterization of the lower envelope $\underline{P}$ on $\mathcal{A}_{\mathcal{L}} \times \mathcal{E}$, starting from a statistical model λ and an unconditional prior belief function Bel obtained as the ϵ-contamination of a reference prior probability. In such case, in (Huber, 1981) it is stated that $\underline{P}(\cdot|E_j)$ is a 2-monotone capacity on $\mathcal{A}_{\mathcal{L}}$, for every $E_j \in \mathcal{E}$, nevertheless, as shown in our Example 2 the envelope $\underline{P}(\cdot|K)$ can fail 2-monotonicity on the whole $\mathcal{A}$, for some $K \in \mathcal{A}^0$.

The following example shows that a full B-conditional prior belief function can be obtained starting from a full conditional prior probability defined on a different algebra.

Example 2 *We consider an automatic system* $\mathbf{S}$ *that can assume three possible states* s_1*,* s_2 *and* s_3*. Let* $\mathcal{S} = \{S_1, S_2, S_3\}$ *be the partition of* Ω*, where* $S_i =$ *"*$\mathbf{S}$ *is in state* s_i*", for* $i = 1, 2, 3$*, and denote* $\mathcal{A}_{\mathcal{S}} = \langle \mathcal{S} \rangle$*. The evolution of* $\mathbf{S}$ *is determined by the Markov chain whose transition matrix and graph are reported in Figure 1.*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

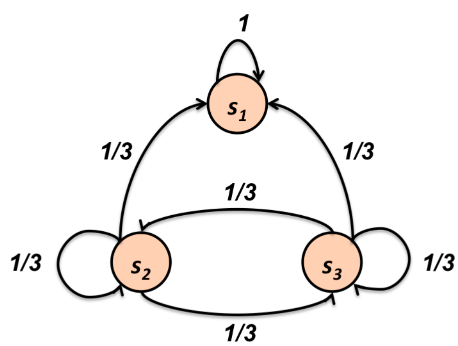

Figure 1: Transition matrix and graph of the Markov chain related to $\mathbf{S}$

Suppose that the initial state of $\mathbf{S}$ *is selected at random and that we observe the system evolve indefinitely in time, so, we can take the limit probabilistic behaviour as our prior information on* $\mathbf{S}$*. The starting probability distribution on the states of* $\mathbf{S}$ *is* $\pi^{(0)} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$*, while after* $n > 0$ *steps we have* $\pi^{(n)} = \pi^{(n-1)} A = \left(1 - \left(\frac{2}{3}\right)^{n+1}, \frac{1}{3}\left(\frac{2}{3}\right)^{n}, \frac{1}{3}\left(\frac{2}{3}\right)^{n}\right)$*.*

It is easily seen that the probability distribution $\pi^{(n)}$ *is positive for every* $n \geq 0$*, so, it uniquely extends to a full conditional probability (still denoted with* $\pi^{(n)}$*) on* $\mathcal{A}_{\mathcal{S}}$ *setting, for every* $A|B \in \mathcal{A}_{\mathcal{S}} \times \mathcal{A}^0_{\mathcal{S}}$*,* $\pi^{(n)}(A|B) = \frac{\pi^{(n)}(A \wedge B)}{\pi^{(n)}(B)}$*. Thus, we have a sequence* $\{\pi^{(n)} \;:\; n = 0, 1, 2, \ldots\}$ *of full conditional probabilities on* $\mathcal{A}_{\mathcal{S}}$ *converging pointwise to the full conditional probability* $\pi^{(\infty)}$ *on* $\mathcal{A}_{\mathcal{S}}$ *defined below*

$\mathcal{A}_{\mathcal{S}}$	$\emptyset$	S_1	S_2	S_3	$S_1 \vee S_2$	$S_1 \vee S_3$	$S_2 \vee S_3$	Ω
$\pi^{(\infty)}(\cdot\|S_1)$	0	1	0	0	1	1	0	1
$\pi^{(\infty)}(\cdot\|S_2)$	0	0	1	0	1	0	1	1
$\pi^{(\infty)}(\cdot\|S_3)$	0	0	0	1	0	1	1	1
$\pi^{(\infty)}(\cdot\|S_1 \vee S_2)$	0	1	0	0	1	1	0	1
$\pi^{(\infty)}(\cdot\|S_1 \vee S_3)$	0	1	0	0	1	1	0	1
$\pi^{(\infty)}(\cdot\|S_2 \vee S_3)$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1
$\pi^{(\infty)}(\cdot\|\Omega)$	0	1	0	0	1	1	0	1

The full conditional probability $\pi^{(\infty)}$ is determined by the complete agreeing class $\{P_0, P_1\}$ of probability measures on $\mathcal{A}_{\mathcal{S}}$ such that $P_0(\cdot) = \pi^{(\infty)}(\cdot|\Omega)$ and $P_1(\cdot) = \pi^{(\infty)}(\cdot|S_2 \vee S_3)$.

Consider now a second automatic system $\mathbf{T}$ that is not directly observable: the only information we have is that $\mathbf{T}$ can assume three possible states t_1, t_2 and t_3, and that if $\mathbf{S}$ is in state s_i then $\mathbf{T}$ is not in state t_i, for $i = 1, 2, 3$. Let $\mathcal{T} = \{T_1, T_2, T_3\}$ be the partition of Ω, where T_i = "$\mathbf{T}$ is in state t_i", for $i = 1, 2, 3$, and denote $\mathcal{A}_{\mathcal{T}} = \langle \mathcal{T} \rangle$ with $T_i \wedge S_i = \emptyset$, for $i = 1, 2, 3$.

As proven in (Coletti et al., 2016b) setting, for every $B \in \mathcal{A}_{\mathcal{T}}$,

$$(B)_* = \bigvee\{S_i \in \mathcal{S} \,:\, S_i \subseteq B\}, \quad Bel_0(B) = P_0((B)_*) \quad \text{and} \quad Bel_1(B) = P_1((B)_*),$$

we obtain a C-class of belief functions $\{Bel_0, Bel_1\}$ on $\mathcal{A}_{\mathcal{T}}$ which, in turn, determines the full B-conditional belief function on $\mathcal{A}_{\mathcal{T}}$ reported below

$\mathcal{A}_{\mathcal{T}}$	$\emptyset$	T_1	T_2	T_3	$T_1 \vee T_2$	$T_1 \vee T_3$	$T_2 \vee T_3$	Ω
$Bel_B(\cdot\|T_1)$	0	1	0	0	1	1	0	1
$Bel_B(\cdot\|T_2)$	0	0	1	0	1	0	1	1
$Bel_B(\cdot\|T_3)$	0	0	0	1	0	1	1	1
$Bel_B(\cdot\|T_1 \vee T_2)$	0	0	0	0	1	0	0	1
$Bel_B(\cdot\|T_1 \vee T_3)$	0	0	0	0	0	1	0	1
$Bel_B(\cdot\|T_2 \vee T_3)$	0	0	0	0	0	0	1	1
$Bel_B(\cdot\|\Omega)$	0	0	0	0	0	0	1	1

Suppose that the state of the unobservable system $\mathbf{T}$ can be verified through a detector $\mathbf{D}$ that returns one of three possible values d_1, d_2 and d_3, with d_i corresponding to the state t_i, for $i = 1, 2, 3$, with a reliability of 90% and equal chances on failures. Let $\mathcal{D} = \{D_1, D_2, D_3\}$ be the partition of Ω, where D_i = "$\mathbf{D}$ returns d_i", for $i = 1, 2, 3$, and denote $\mathcal{A}_{\mathcal{D}} = \langle \mathcal{D} \rangle$. Let $\mathcal{A} = \langle \mathcal{A}_{\mathcal{T}} \cup \mathcal{A}_{\mathcal{D}} \rangle$ and consider the statistical model on $\mathcal{A}_{\mathcal{D}} \times \mathcal{T}$ singled out by

$$\lambda(D_i|T_i) = 90\%, \quad \lambda(D_j|T_i) = \lambda(D_k|T_i) = 5\%, \quad \text{for different } i, j, k \in \{1, 2, 3\},$$

that uniquely extends to a strategy σ on $\mathcal{A} \times \mathcal{T}$ by Proposition 1 in (Petturiti and Vantaggi, 2017).

The full B-conditional belief function Bel_B encodes all our prior information on $\mathbf{T}$ and can be used together with σ to draw Bayesian inferences. At this aim, suppose that the detector $\mathbf{D}$ shows the value d_j, for $j = 1, 2, 3$, then the lower posterior distribution on the states of $\mathbf{T}$ can be easily determined using Proposition 8. For instance, since $\underline{P}(D_j|\Omega) > 0$, $\underline{P}(T_1 \wedge D_j|\Omega) = 0$ and $\overline{P}(T_1^c \wedge D_j|\Omega) > 0$, for $j = 1, 2, 3$, we get

$$\underline{P}(T_1|D_j) = \frac{\underline{P}(T_1 \wedge D_j|\Omega)}{\underline{P}(T_1 \wedge D_j|\Omega) + \overline{P}(T_1^c \wedge D_j|\Omega)} = 0,$$

and, analogously, we can compute $\underline{P}(T_1^c|D_j) = 1$, so, $\underline{P}(T_1|D_j) = \overline{P}(T_1|D_j) = 0$, i.e., the observation of the detector $\mathbf{D}$ does not change our degree of belief on T_1 since it is $Bel_B(T_1|\Omega) = Pl_B(T_1|\Omega) = 0$.

5. Conclusions

We show that, as long as we consider a precise strategy σ, the introduction of ambiguity in the prior information through a full B-conditional belief function Bel_B has straightforward treatment:

a characterization for the envelopes of the class of full conditional probabilities dominating the assessment $\{Bel_B, \sigma\}$ is provided. The entire procedure lives inside Williams framework and the characterized lower envelope reveals to be the natural extension of $\{Bel_B, \sigma\}$. Our aim for future research is to introduce ambiguity also in the strategy by considering an imprecise strategy β such that $\beta(\cdot|H_i)$ is a belief function, for every $H_i \in \mathcal{L}$, possibly removing the finiteness assumption. This would lead to a theory to compare with that of (Walley, 1991).

Acknowledgments

This work was partially supported by INdAM-GNAMPA through the Project 2016 `U2016/000391`.

References

A. Capotorti, L. Galli, and B. Vantaggi. Locally strong coherence and inference with lower–upper probabilities. *Soft Computing*, 7(5):280–287, 2003.

G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.

G. Coletti and R. Scozzafava. *Probabilistic Logic in a Coherent Setting*, volume 15 of *Trends in Logic*. Kluwer Academic Publisher, Dordrecht/Boston/London, 2002.

G. Coletti, D. Petturiti, and B. Vantaggi. Bayesian inference: the role of coherence to deal with a prior belief function. *Statistical Methods and Applications*, 23(4):519–545, 2014.

G. Coletti, D. Petturiti, and B. Vantaggi. When upper conditional probabilities are conditional possibility measures. *Fuzzy Sets and Systems*, 304:45–64, 2016a.

G. Coletti, D. Petturiti, and B. Vantaggi. Conditional belief functions as lower envelopes of conditional probabilities in a finite setting. *Information Sciences*, 339:64–84, 2016b.

B. de Finetti. *Theory of Probability 1-2*. Wiley, 1975.

A. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

D. Denneberg. *Non-Additive Measure and Integral*, volume 27 of *Series B: Mathematical and Statistical Methods*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1994.

L. DeRoberts and J. Hartigan. Bayesian inference using intervals of measures. *Annals of Statistics*, 9(2):235–464, 1981.

L. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3(1):89–99, 1975.

D. Dubois and T. Denœux. *Conditioning in Dempster-Shafer Theory: Prediction vs. Revision*, pages 385–392. Springer Berlin Heidelberg, 2012.

R. Fagin and J. Halpern. *Uncertainty in Artificial Intelligence*, chapter A New Approach to Updating Beliefs, pages 347–374. Elsevier Science Publishers, 1991.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

S. Holzer. Sulla nozione di coerenza per le probabilità subordinate. In *Rendiconti dell'Istituto di Matematica dell'Università di Trieste*, volume 16, pages 46–62. 1984.

P. Huber. *Robust Statistics*. Wiley, 1981.

J.-Y. Jaffray. Bayesian updating and belief functions. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(5):1144–1152, 1992.

P. Krauss. Representation of conditional probability measures on boolean algebras. *Acta Mathematica Academiae Scientiarum Hungarica*, 19(3-4):229–241, 1968.

D. Petturiti and B. Vantaggi. Envelopes of conditional probabilities extending a strategy and a prior probability. *International Journal of Approximate Reasoning*, 81:160–182, 2017.

E. Regazzini. Finitely additive conditional probabilities. *Rendiconti del Seminario Matematico e Fisico di Milano*, 55(1):69–89, 1985.

D. Schmeidler. Integral representation without additivity. *Proceedings of the American Mathematical Society*, 97:255–261, 1986.

G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.

E. Torgersen. *Comparison of Statistical Experiments*, volume 36 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1991.

M. Troffaes and G. de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. Wiley, 2014.

P. Walley. Coherent lower (and upper) probabilities. Technical report, Department of Statistics, University of Warwick, 1981.

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

L. Wasserman. Prior envelopes based on belief functions. *Annals of Statistics*, 18(1):454–464, 1990a.

L. Wasserman. Belief functions and statistical inference. *Canadian Journal of Statistics*, 18(3): 183–196, 1990b.

L. Wasserman and J. Kadane. Bayes' theorem for choquet capacities. *Annals of Statistics*, 18(3): 1328–1339, 1990.

P. Williams. Note on conditional previsions. Unpublished report of School of Mathematical and Physical Science, University of Sussex (Published in *International Journal of Approximate Reasoning*, 44:366–383, 2007), 1975.