Probabilistic argumentation systems
A new way to combine logic with probability

Jürg Kohlas

Department of Informatics DIUF, University of Fribourg, CH-1700 Fribourg, Switzerland

Abstract

Probability is usually closely related to Boolean structures, i.e., Boolean algebras or propositional logic. Here we show, how probability can be combined with non-Boolean structures, and in particular non-Boolean logics. The basic idea is to describe uncertainty by (Boolean) assumptions, which may or may not be valid. The uncertain information depends then on these uncertain assumptions, scenarios or interpretations. We propose to describe information in information systems, as introduced by Scott into domain theory. This captures a wide range of systems of practical importance such as many propositional logics, first order logic, systems of linear equations, inequalities, etc. It covers thus both symbolic as well as numerical systems. Assumption-based reasoning allows then to deduce supporting arguments for hypotheses. A probability structure imposed on the assumptions permits to quantify the reliability of these supporting arguments and thus to introduce degrees of support for hypotheses. Information systems and related information algebras are formally introduced and studied in this paper as the basic structures for assumption-based reasoning. The probability structure is then formally represented by random variables with values in information algebras. Since these are in general non-Boolean structures some care must be exercised in order to introduce these random variables. It is shown that this theory leads to an extension of Dempster–Shafer theory of evidence and that information algebras provide in fact a natural frame for this theory.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Logic; Probability; Assumption-based reasoning; Inference; Uncertainty; Evidence theory; Dempster–Shafer theory

* Research supported by grant No. 2000-061454 of the Swiss National Foundation for Research.
E-mail address: juerg.kohlas@unifr.ch (J. Kohlas).
URL: http://diuf.unifr.ch/
1. Introduction

Argumentation has gained growing recognition as a new and promising research direction in reasoning, and in particular, reasoning under uncertainty. Different authors have investigated argumentation and its application in various domains. By looking at today’s literature on this subject, one realizes that argumentation is understood in fairly different ways. The variety of attempts to study the nature of arguments and the process of argumentation is therefore characterized by a broad diversity.

It is not possible to give here a comprehensive account of what argumentation means in its different interpretations, nor is it possible to list all relevant references to this subject. Therefore only some work which has some connections to the approach presented here can be mentioned. Many authors focus on the problem of the acceptability and the comparison of arguments, e.g., [1,8]. A recent contribution in this area is the logic-based theory of deductive arguments [4]. Some authors consider argumentation as a dialectical process during disputations [28]. Other authors speak about negotiation [21]. We refer to [22] for a fairly recent and comprehensive overview of modeling of argumentation. The common feature of all these papers is their restriction to purely logical analysis. It is characteristic of this work that it attempts to avoid the use of probabilities to compare or select relevant arguments. This is in contrast to the approach presented in this paper, where it is proposed to combine probability with logic in a new way.

The basic idea of the probabilistic argumentation systems framework goes back to the concept of assumption-based truth maintenance systems (ATMS) [6]. The goal is not to describe argumentation as a dialectic process, but rather to serve as a deductive tool that helps to judge hypotheses, i.e., open questions about the unknown state of a past, present or future world. This will be done using the available uncertain, partial and possibly even contradictory knowledge and information. From a qualitative point of view, the problem is to derive arguments in favor of hypotheses of interest. An argument is a defeasible proof built on certain assumptions, i.e., a deduction of a hypothesis based on assumptions. For this, the knowledge and the information must be represented in the language of an appropriate logic which provides the necessary deduction machinery. This can be propositional or first order logic. It can also be some more restricted system such as systems of linear equations or inequalities.

Although a qualitative judgment of a hypothesis, based on arguments, may be valuable, a quantitative measurement of the degree of credibility or support of the hypothesis may be more useful and help to decide whether a hypothesis can be believed in or not. For this purpose a probabilistic structure can be defined on top of argumentation systems. It will permit to compute the reliability of arguments and thus to measure the degree of support of hypotheses.

This is the sketch of a novel way to combine logic and probability, the two classic tools of inference. Logic serves to deduce arguments, probability to weigh them by their reliability. We shall use classical, monotonic logic. Basically first order logic would be an appropriate choice. However, this system is, especially for computations, often too complex. Therefore, we shall use another abstract framework, information systems, for our presentation. This will cover all practical systems such as propositional logic, finite constraint systems, systems of linear equations, and many others, for which efficient
algorithms to construct arguments exist. Information systems and their use to define argumentation systems will be described in Section 2. In this section we shall also introduce the basic concepts of argumentation systems as understood in this paper.

Abstract argumentation systems induce in a natural way an interesting algebraic structure of assumption-based reasoning. In Section 3 this algebraic background of the theory will be developed. It will be shown how information systems give rise to information algebras. These algebras permit to describe the combination of pieces of information, and also to define a partial order of information content. We use here a simplified version of an algebra introduced in [13]. This will be sufficient for our purpose. Uncertainty is represented by a mapping from assumptions into an information algebra. This setting permits to define a probabilistic structure on top of it in a natural way. The argumentation system induces then a random variable with values in the associated information algebra. Since these algebras are non-Boolean some care must be taken in defining random variables. The approach proposed here constitutes one of the major new contributions of this paper. This probabilistic structure will be examined in Section 4. It will be shown that it corresponds to Dempster–Shafer theory [7,25] in an abstract setting.

The way we propose to combine logic and probability in this paper has already been proposed in [17] in the particular context of propositional logic. It has been worked out in this context to some extent in [9,14]. Especially the computational problems have been examined in some depth and a software system for propositional probabilistic argumentation ABEL has been developed [2]. On the other hand [19] adopted the generic approach described in this paper for the kind of numerical systems found in statistical inference, especially for linear systems with Gaussian disturbances. Information algebras in a somewhat richer version have been introduced in [13]. Random variables in information algebras have also been introduced in [13]. But the theory presented here is more complete and contains new results.

2. Abstract argumentation systems

In order to specify information, we need an appropriate language, together with a means to deduce consequences of the information. Therefore, we assume a language \( \mathcal{L} \), which, for our purposes, can be considered simply as a set of tokens or sentences. We do not need to worry about the syntactical structure of sentences at this point. Information, i.e., data and facts are then simply given as subsets \( X \subseteq \mathcal{L} \). The idea that certain other sentences can be deduced from \( X \) is represented by an entailment relation defined between subsets of \( X \subseteq \mathcal{L} \) and single sentences \( s \in \mathcal{L} \). The notation \( X \models s \) says that the sentence \( s \) is entailed by the set of sentences \( X \), or \( X \) entails \( s \). This relation must satisfy the following conditions:

\begin{align*}
(E1) \quad & X \models s \text{ for all } s \in X, \\
(E2) \quad & X \models b \text{ for all } b \in Y \text{ and } Y \models a \text{ imply } X \models a.
\end{align*}

These are the usual conditions required for an entailment relation. The system \( (\mathcal{L}, \models) \) is called an information system. The notion of an information system has been introduced
by [23], but we use it here in a slightly different way; in particular, we do not single out finite consistent sets of tokens.

We sketch here a few nontrivial examples of information systems.

(1) **Systems of linear equations.** Consider a finite set of variables $X_1, \ldots, X_n$ and a field $\mathcal{F}$. Let $\mathcal{L}$ be the set of linear equations of the form

$$a_0 + a_1X_1 + \cdots + a_nX_n = 0,$$

with coefficients $a_i \in \mathcal{F}$. The entailment relation is defined using linear combination: Let $X \subseteq \mathcal{L}$, $l \in \mathcal{L}$, then $X \vdash l$, if there is finite subset $Y = \{l_1, \ldots, l_m\} \subseteq X$ such that

$$l = \lambda_1 l_1 + \cdots + \lambda_m l_m,$$

with $\lambda_j \in \mathcal{F}$. We remark, that a similar system can be defined for linear inequalities over ordered fields, where entailment is defined by positive linear combination.

(2) **Propositional logic.** Propositional logic is an information system. In fact, it is an example of the following more general system.

(3) **Contexts.** A context is given by two sets $M$ (whose elements are called models) and $S$ (whose elements are called sentences) and a relation $\models \subseteq M \times S$. We write $m \models s$ ($m$ models $s$) for $(m, s) \in \models$. We define for any subset $B \subseteq S$ the extend of $B$ \[ \hat{r}(B) = \{m \in M: m \models s \text{ for all } s \in B \}. \]

Next, we say the relation $B \vdash s$ holds, if $\hat{r}(B) \subseteq \hat{r}(\{s\})$. It is easy to verify that $\vdash$ is an entailment relation. So, $(S, \vdash)$ is an information system related to the context $(M, S, \models)$. This covers many logical systems: The elements of $M$ are the structures of a logical language $S$, which serve to evaluate sentences of $S$ to true or false. Contexts or Chu-spaces are discussed in [3,5]. We refer to [18] for a discussion of a similar structure in a similar spirit as here, and also for several examples of logics falling into this framework.

If $(\mathcal{L}, \vdash)$ is an information system, we introduce the operator

$$C(X) = \{s \in \mathcal{L}: X \vdash s\}.$$ 

$C$ is a mapping $C : \mathcal{P}(\mathcal{L}) \to \mathcal{P}(\mathcal{L})$. It is well known, and easy to verify that $C$ has the following properties:

1. $X \subseteq C(X)$ for all $X \subseteq \mathcal{L}$.
2. $C(C(X)) = C(X)$ for all $X \subseteq \mathcal{L}$.
3. $Y \subseteq X \subseteq \mathcal{L}$ implies $C(Y) \subseteq C(X)$.

Thus, $C$ is a consequence operator (or also closure operator). $C(X)$ is called the closure of $X$. A set $X$ such that $X = C(X)$ is called closed. And we say that $Y$ is a consequence of $X$, written as $X \vdash Y$, if $Y \subseteq C(X)$. Two sets $X$ and $Y$ are called equivalent, if $C(X) = C(Y)$. 
Uncertainty is introduced into an information system by noting that an information may depend on uncertain assumptions, on uncertain scenarios or may have an uncertain interpretation. This is captured by a set $\Omega$ whose elements $\omega$ represent possible assumptions, scenarios or interpretations on which depends a given body of knowledge or information. If a scenario $\omega \in \Omega$ is assumed to hold, then we suppose that the information is given by $X(\omega) \in \mathcal{L}$. The mapping 

$$X : \Omega \rightarrow \mathcal{P}(\mathcal{L})$$

which assigns to each possible assumption, scenario or interpretation $\omega$ the corresponding facts expressed by $X(\omega)$ is called an assumption-based information.

We admit that the possible assumptions $\Omega$ are specified before the assumption-based information $X$ becomes available. This implies that this information may render some originally possible interpretations $\omega$ impossible. These are the interpretations, which are in contradiction to $X$, that is those for which $X(\omega) = \mathcal{L}$. Therefore, we define the set of inconsistent scenarios by

$$IX = \{ \omega \in \Omega : C(X(\omega)) = \mathcal{L} \}. \quad (2.1)$$

$CX = \Omega \setminus IX$ are then the consistent scenarios (relative to $X$).

We may ask whether some hypothesis, expressed by a sentence $s \in \mathcal{L}$ or more generally by some subset $H \subseteq \mathcal{L}$ could possibly be true in the light of the given information. This will depend on the scenario or the interpretation we assume. Under a scenario $\omega$ the hypothesis $H$ is necessarily valid, if $X(\omega) \vdash H$. We then consider the set of scenarios under which $H$ would be necessarily true (given the information): This is the set of all consistent scenarios, which imply $H$:

$$sX(H) = \{ \omega \in CX : X(\omega) \vdash H \} = \{ \omega \in CX : H \subseteq C(X(\omega)) \}. \quad (2.2)$$

This is called the support for $H$. For technical reasons, it is convenient to introduce the so-called quasi-support, which contains all scenarios, including the inconsistent ones, which support $H$:

$$qsX(H) = \{ \omega \in \Omega : X(\omega) \vdash H \} = \{ \omega \in \Omega : H \subseteq C(X(\omega)) \}.$$  

Then we see that $IX = qsX(\mathcal{L})$ and $sX(H) = qsX(H) \setminus qsX(\mathcal{L})$. These concepts will be important for the probabilistic structure introduced in Section 4 later. The structure presented so far represents the logical part of our model of uncertain information.

To illustrate these basic ideas we present two small examples:

1. **Sensor.** A sensor should detect some alarm state (like smoke in a computer center). It may fail in two ways: It may be unable to react to an alarm state or it may wrongly give an alarm. We model this in propositional logic as follows: We define the propositional variables

- $h$ alarm state (to be detected),
- $e$ alarm (evidence for alarm state),
and the assumptonal variables

\begin{align*}
ok1 & \quad \text{sensor is ok to detect alarm state,} \\
ok2 & \quad \text{sensor is ok not to signal not existing alarm state.}
\end{align*}

We describe the sensor system by the following logical statements:

\begin{align*}
h & \land ok1 \rightarrow e, \\
\neg ok1 & \rightarrow \neg e, \\
\neg h & \land ok2 \rightarrow \neg e, \\
\neg ok2 & \rightarrow e.
\end{align*}

Suppose that the alarm rings. Then we add the statement

\begin{equation*}
e.
\end{equation*}

We may then verify that the only consistent scenarios are

\begin{equation*}
\{ \{ok1, ok2\}, \{ok1, \neg ok2\} \}.
\end{equation*}

The scenarios simply say that \(ok1\) must be true. The single supporting scenario for the hypothesis \(h\) that an alarm state is present is \(\{ok1, ok2\}\). There is no supporting scenario for the hypothesis \(\neg h\) that there is no alarm state despite the alarm ringing. We refer to [9] for methods to compute such sets of consistent and supporting scenarios.

2. Noisy communication. When a signal \(X\) is transmitted through a communication channel, then noise may disturb it, so that the output corresponds not exactly to the input. Therefore, we may transmit the signal through two parallel channels, to be in a better position to reconstruct the input from the output. We model this in a numerical frame introducing the real-valued variables

\begin{align*}
X & \quad \text{input signal,} \\
Y_1, Y_2 & \quad \text{output signals of the two channels,}
\end{align*}

and further the disturbances

\begin{align*}
\omega_1, \omega_2 & \quad \text{noise in the two channels.}
\end{align*}

The transmission of the input signal over the two channels is then determined by

\begin{align*}
Y_1 & = X + \omega_1, \\
Y_2 & = X + \omega_2.
\end{align*}

Suppose that we actually observe the values \(y_1\) and \(y_2\) for the two output signals. Then we add the equations

\begin{align*}
Y_1 & = y_1, \\
Y_2 & = y_2.
\end{align*}
Then we easily find out the only consistent scenarios \((\omega_1, \omega_2)\) are those, which satisfy the equation

\[
\omega_1 - \omega_2 = y_1 - y_2.
\]

In order to reconstruct the input signal from the output signals we are interested in supporting scenarios for hypotheses like \(X = x\) for some specified values \(x\). It is easy to see that there is a single supporting scenario for this hypothesis:

\[
s\left(\{X = x\}\right) = \{ (\omega_1, \omega_2): \omega_1 = y_1 - x, \ \omega_2 = y_2 - x \}.
\]

Such linear models with disturbances and methods for their computational treatment are discussed in [19].

We may say that the larger the support set of a hypothesis, the more the hypothesis is credible in the light of the information. But clearly this is in many cases insufficient to measure the credibility of a hypothesis. Different scenarios, assumptions or interpretations may have different likelihood. This is where probability enters. It serves to express the likelihood of scenarios and to compute then the degree of support of hypotheses based on these probabilities. In order to discuss this probabilistic structure, we need first to study the algebraic structure of the concept of uncertain information introduced.

3. The algebraic structure of argumentation

3.1. Information algebras

If \((\mathcal{L}, \vdash)\) is an information system, we have said that \(X \subseteq \mathcal{L}\) is an information. But what is the information given by \(X\) really? Note that we may deduce further sentences \(Y \subseteq C(X)\) from \(X\). So \(Y\) belongs also to the information. In fact, the closure \(C(X)\) is the whole information specified by \(X\) and \(X\) is only one possible representation of it. Any equivalent set \(Y\) with \(C(Y) = C(X)\) would be another representation.

Therefore any closed set \(E \subseteq \mathcal{L}\) (i.e., any set such that \(E = C(E)\)) is taken to be an information. And any set \(X\) such that \(C(X) = E\) is a representation of information \(E\). We denote by \(\mathcal{L}_C\) the set of all closed sets in \(\mathcal{L}\). Note that \(\mathcal{L} \in \mathcal{L}_C\) and \(C(\emptyset) \in \mathcal{L}_C\). The former is called the contradiction or null information, the latter the vacuous or neutral information.

We note that information can be combined. Informally, if \(X_1 \subseteq \mathcal{L}\) and \(X_2 \subseteq \mathcal{L}\) represent each some piece of information, then together \(X_1 \cup X_2\) represent the combined or aggregated information. The corresponding pieces of information are \(C(X_1), C(X_2)\) and \(C(X_1 \cup X_2)\). We remark that \(C(X_1 \cup X_2) = C(C(X_1) \cup C(X_2))\). Therefore, we define an operation \(\otimes\) of combination between information in \(\mathcal{L}_C\) by

\[
E_1 \otimes E_2 = C(E_1 \cup E_2).
\]

Clearly this operation is commutative and associative, i.e., \(\mathcal{L}_C\) is a commutative semigroup under the operation \(\otimes\). The vacuous information is the neutral element of this semigroup,

\[
E \otimes C(\emptyset) = C(E \cup \emptyset) = C(E).
\]
The null information \( \mathcal{L} \) is absorbing in the semigroup
\[
\mathcal{E} \otimes \mathcal{L} = \mathcal{E} \otimes \mathcal{C}(\mathcal{L}) = \mathcal{C}(\mathcal{E} \cup \mathcal{L}) = \mathcal{C}(\mathcal{L}) = \mathcal{L}.
\]
Finally, the semigroup is idempotent,
\[
\mathcal{E} \otimes \mathcal{E} = \mathcal{C}(\mathcal{E} \cup \mathcal{E}) = \mathcal{C}(\mathcal{E}) = \mathcal{E}.
\]
We call a commutative and idempotent semigroup (i.e., a semilattice, see below) with a neutral and a null element an \textit{information algebra}.\(^1\)

In a more abstract setting, we may therefore consider an information algebra \( \Phi \), whose generic elements will be called \( \phi, \psi, \ldots \), whose neutral element is \( e \) and whose null element is denoted by \( z \). Since \( \Phi \) is idempotent, we may define a partial order between elements of \( \Phi \): \( \phi \leq \psi \) if, and only if \( \phi \otimes \psi = \psi \). An information \( \phi \) is less informative than \( \psi \) if combined with the latter, nothing new results. This is indeed a partial order, as is easily verified. We have \( e \leq \phi \leq z \) for all \( \phi \in \Phi \). And in particular, we have
\[
\phi \otimes \psi = \phi \lor \psi.
\]
In fact, \( \phi, \psi \leq \phi \otimes \psi \). Assume another upper bound \( \eta \) of \( \phi \) and \( \psi \). Then
\[
(\phi \otimes \psi) \otimes \eta = \phi \otimes (\psi \otimes \eta) = \phi \otimes \eta = \eta,
\]
hence \( \phi \otimes \psi \leq \eta \). Henceforth we shall at convenience replace combination by the supremum (the join in the semilattice). Therefore, \( \Phi \) is in fact a \textit{semilattice}. In the example induced by an information system for example, \( E_1 \leq E_2 \) means that \( E_1 \subseteq E_2 \). Note that the semilattice induced by an information system is even complete: For any family of closed sets \( E_i, i \in I \), we have
\[
\mathcal{E} = \mathcal{C}\left( \bigcup_{i \in I} E_i \right) = \bigvee_{i \in I} E_i.
\]
Indeed, assume \( F \supseteq \mathcal{E} \). Then \( F \supseteq E_i \) for all \( i \in I \), hence \( F \supseteq \bigcup E_i \) and \( F = \mathcal{C}(F) \supseteq \mathcal{C}(\bigcup E_i) = \mathcal{E} \). So, \( \mathcal{C}(\bigcup E_i) \) is the supremum of the \( E_i \). We do not assume however in general that an information algebra is a complete semilattice.

Any information system induces an information algebra, as shown above. Let us illustrate this with the examples of the previous Section 2.

1. \textit{Contexts}. Consider a context \( (M, S, \models) \). For \( A \subseteq M \) we define the intend of \( A \)
\[
\hat{\tau}(A) = \{ s \in S: m \models s \text{ for all } m \in A \}.
\]
Then, for a \( B \subseteq S \),
\[
\mathcal{C}(B) = \{ s \in S: B \vdash s \} = \hat{\tau}(\hat{\tau}(B)).
\]

\(^1\) There is an important element missing: information relates to questions. Therefore an algebra of information should model systems of related questions and contain an operation of focusing of information to particular questions. This leads to a much richer algebraic structure [13,16]. Here we restrict ourselves, for the sake of simplicity, to the case of one single fixed question.
We refer to [12] for a proof and more details. In a similar way we define a consequence operator in $M$:

$$C(A) = \hat{r}(\hat{r}(A)).$$

An information in the context $(M, S, \models)$ is then a pair $(A, B)$ such that $A = \hat{r}(B)$ and $B = \hat{r}(A)$. Clearly, both $A$ and $B$ are closed sets. What we call here an information is also called a concept in concept analysis [5].

2. Propositional logic. This system can be seen as a context (see above), where $S$ is the propositional language and $M$ the set of truth functions. $m \models s$ means that $s$ evaluates to true under $m$. In this particular context all subsets of $M$ are closed. $B \models s$ holds, if all models $\hat{r}(B)$ are also models of $s$, i.e., if $\hat{r}(B) \subseteq \hat{r}([s])$. So $C(B) = \hat{r}(A)$ and any pair $(A, \hat{r}(A))$ represents an information. We can also determine a propositional information from a set $B$ of formulas by $(\hat{r}(B), C(B))$. In terms of models, combination is simply intersection,

$$(\hat{r}(A_1), \hat{r}(A_1)) \otimes (\hat{r}(A_2), \hat{r}(A_2)) = (A_1 \cap A_2, \hat{r}(A_1 \cap A_2)).$$

3. Linear equations. Here we have again a context, $M$ is the vector space $F^n$, $S$ the set of linear equations over $n$ variables with coefficients in $F$. For $x \in F^n$ and $l \in S$, the relation $x \models l$ means that $x$ is a solution of $l$. For a system of linear equations $B \subseteq S$ the set $\hat{r}(B)$ is the linear solution manifold of the system of equations. $C(B)$ is the set of all linear equations whose solution manifold contains the solution manifold of $B$. Conversely, for any subset $A \in F^n$, the closure $C(A)$ is the linear manifold spanned by the elements of $A$. Combination in this context amounts to intersect linear manifolds.

Here follow a few further simple examples of information algebras, which are not derived from information systems.

1. Subsets. Subsets of a reference set $S$ form an information algebra with intersection for combination. The set $S$ is the neutral element of this algebra and the emptyset the null element. We remark that this system has a Boolean structure (it is a distributive lattice with complement), as have many other, but not all information algebras.

2. Binary strings. Let $\{0, 1\}^n$ be the set of all finite binary strings. For $x, y \in \{0, 1\}^n$ we say that $x \leq y$ if $x$ is a prefix of $y$ (i.e., an initial substring of $y$), $\varepsilon$ is the empty string, $\varepsilon \leq x$ for all $x \in \{0, 1\}^n$. We adjoin an element $z$ to $\{0, 1\}^n$ and put $\Phi = \{0, 1\}^n \cup \{z\}$. We define a combination in $\Phi$ as follows: $x \otimes y = y$, if $x \leq y$, $x \otimes y = z$ if neither $x \leq y$ nor $y \leq x$ (we say that $x$ and $y$ are not consistent) and $x \otimes z = z$ for all $x$. This is an information algebra, albeit with a somewhat trivial combination operation.

3. Intervals. Intervals $[x, y]$ with $-\infty \leq x \leq y \leq +\infty$ form an information algebra with intersection for combination, if we add the empty interval as null element.

An ideal in an information algebra $\Phi$ is a subset $I \subseteq \Phi$, such that

1. $\psi \leq \phi \in I$ implies $\psi \in I$.
2. $\phi, \psi \in I$ implies $\phi \otimes \psi \in I$.

The set $\{\psi \in \Phi : \psi \leq \phi\}$ is an ideal $I(\phi)$, a so-called principal ideal. Note that every ideal contains $e$. An ideal different from $\Phi$ is called proper. We may combine ideals by

$$I_1 \otimes I_2 = \{\phi \in \Phi : \phi \leq \phi_1 \otimes \phi_2 \text{ for some } \phi_1 \in I_1, \phi_2 \in I_2\}.$$
This is clearly an ideal. This operation is commutative and associative and the ideal $I(e)$ is the neutral element, $\Phi$ the null element. Thus the set of ideals $I(\Phi)$ of an information algebra $\Phi$ is itself an information algebra. The mapping $\phi \mapsto I(\phi)$ is an embedding, since $e \mapsto I(e)$ and $I(\epsilon) = \Phi$, $I(\phi) = I(\psi)$ implies $\phi = \psi$ and clearly

$$I(\phi \otimes \psi) = I(\phi) \otimes I(\psi).$$

Therefore, we consider $I(\Phi)$ as an extension of $\Phi$, its completion. Note then that in $I(\Phi)$ we have $\phi \leq I$ if, and only if, $\phi \in I$.

**Lemma 1.** For all $I_1, I_2 \in I(\Phi)$,

$$I_1 \otimes I_2 = \bigvee \{ \phi_1 \otimes \phi_2 : \phi_1 \in I_1, \phi_2 \in I_2 \}.$$  \hfill (3.1)

**Proof.** For all $\phi \in I_1 \otimes I_2$ we have by definition of $I_1 \otimes I_2$ a $\phi_1 \in I_1$ and a $\phi_2 \in I_2$ such that $\phi \leq \phi_1 \otimes \phi_2$. Therefore

$$I_1 \otimes I_2 \leq \bigvee \{ \phi_1 \otimes \phi_2 : \phi_1 \in I_1, \phi_2 \in I_2 \}.$$  But

$$\{ \phi_1 \otimes \phi_2 : \phi_1 \in I_1, \phi_2 \in I_2 \} \subseteq \{ \phi \in \Phi : \phi \leq \phi_1 \otimes \phi_2 \text{ for some } \phi_1 \in I_1, \phi_2 \in I_2 \}.$$  Hence we see that

$$I_1 \otimes I_2 \geq \bigvee \{ \phi_1 \otimes \phi_2 : \phi_1 \in I_1, \phi_2 \in I_2 \}.$$  This proves the identity (3.1). \quad \square

We show next that to any information algebra $\Phi$, we can associate an information system. In fact let $\mathcal{L} = \Phi$. We define an entailment relation $\vdash_\Phi$ for $X \subseteq \Phi$ and $\phi \in \Phi$ as follows:

$$X \vdash_\Phi \phi \quad \text{if, and only if, there is } F \subseteq X, \text{ finite such that } \phi \leq \bigotimes_{\psi \in F} \psi.$$  \hfill (3.2)

This is indeed an entailment relation as can easily be verified. Therefore, we have an information system $(\Phi, \vdash_\Phi)$ induced by an information algebra $\Phi$. Its associated consequence operator is denoted by $C_\Phi$. We remark that $C_\Phi(\emptyset) = I(e)$, since we interpret the combination of an empty family of information as $e$. Further we note that $C_\Phi(X)$ is an ideal in $\Phi$. We call this ideal $I(X)$, the ideal generated by a subset $X$ of $\Phi$. So, the closed sets of this information systems are exactly the ideals of $\Phi$. As we know this is an information algebra $I(\Phi)$. The following theorem shows that it is also the information algebra induced by the information system $(\Phi, \vdash_\Phi)$.

**Theorem 1** [13]. If $\Phi$ is an information algebra, we have, for $X_1, X_2 \subseteq \Phi$

$$C_\Phi(X_1 \cup X_2) = I(X_1) \otimes I(X_2).$$
Proof. Suppose $\phi \in C_{\Phi}(X_1 \cup X_2)$. Then there is a finite set $F \subseteq X_1 \cup X_2$ such that $\phi \leq \bigotimes_{\psi \in F} \psi$. Let $F_1 = F \cap X_1$ and $F_2 = F \cap X_2$. Then $F_1$ and $F_2$ are finite and subsets of $X_1$ and $X_2$ respectively. Define

$$\psi_1 = \bigotimes_{\psi \in F_1} \psi, \quad \psi_2 = \bigotimes_{\psi \in F_2} \psi.$$ 

Then we have $\psi_1 \in I(X_1)$, $\psi_2 \in I(X_2)$ and $\phi \leq \psi_1 \otimes \psi_2$, hence $\phi \in I(X_1) \otimes I(X_2)$.

Conversely, if $\phi \in I(X_1) \otimes I(X_2)$, then $\phi \leq \psi_1 \otimes \psi_2$ for $\psi_1 \in I(X_1)$ and $\psi_2 \in I(X_2)$. There are then finite sets $F_1 \subseteq X_1$ and $F_2 \subseteq X_2$ such that

$$\psi_1 \leq \bigotimes_{\psi \in F_1} \psi, \quad \psi_2 \leq \bigotimes_{\psi \in F_2} \psi.$$ 

But then it follows that

$$\phi \leq \bigotimes_{\psi \in F_1 \cup F_2} \psi$$

and $F_1 \cup F_2 \subseteq X_1 \cup X_2$, finite. Therefore, $\phi \in C_{\Phi}(X_1 \cup X_2)$. \hfill \Box

According to this theorem, the information algebra induced by the information system $(\Phi, \vdash_{\Phi})$ is the completion $I(\Phi)$ of the algebra $\Phi$. Hence information systems and information algebras are two sides of the same coin and we may work with information systems or information algebras at our convenience.

3.2. Compact systems

If we express information in an information system $(L, \vdash)$, then we can effectively only express information by finite sets $X \subseteq L$. However, we would want to express any information at least approximatively by an effective information. This is achieved in a compact information system, where the following additional property of the entailment relation is satisfied:

If $X \vdash s$, then there is a finite subset $F \subseteq X$ such that $F \vdash s$.

The corresponding consequence operator $C$ has then the following additional property as is well-known [5]

$$C(X) = \bigcup \{ C(F) : F \subseteq X, \text{ } F \text{ finite} \}.$$ 

Such a consequence operator is called compact too. The closed sets $C(X)$, where $X$ is finite, i.e., finitely axiomatizable sets, can be considered as the information which is “finite” or effective. All other information can be approximated by the union above, because if $F_1, F_2 \subseteq X$ are finite sets, then $F_1 \cup F_2 \subseteq X$ is a finite set, which gives a better approximation.

We claim that this compact information algebra $L_C$ is an instance of the following algebraic structure: Suppose $\Phi$ an information algebra. Then a subset $D \subseteq \Phi$ is called directed, if for $\phi, \psi \in D$ there is an upper bound $\eta \in D$ such that $\phi, \psi \leq \eta$. Assume $\Phi_f \subseteq \Phi$ a subset of “finite” elements such that the following is satisfied:

"
(1) Finitary semigroup. \( \phi, \psi \in \Phi_f \) implies \( \phi \otimes \psi \in \Phi_f \), and \( e \in \Phi_f \).

(2) Convergence. If \( D \subseteq \Phi_f \) is a directed set, then \( \bigvee D \) exists and belongs to \( \Phi \).

(3) Density. For all \( \phi \in \Phi \),

\[
\phi = \bigvee \{ \psi \in \Phi_f : \psi \leq \phi \}.
\]

(4) Compactness. If \( D \subseteq \Phi_f \) is a directed set, and \( \phi \in \Phi_f \) such that \( \phi \leq \bigvee D \), then there is a \( \psi \in D \) such that \( \phi \leq \psi \).

Such an information algebra \((\Phi, \Phi_f)\) is called finitary or compact. The following theorem is proved in [13,16]:

**Theorem 2.** Let \((\Phi, \Phi_f)\) be a finitary information algebra. Then

1. \( \Phi \) is a complete lattice.
2. \( \phi \in \Phi_f \) if, and only if, for all directed subsets \( D \subseteq \Phi \), \( \phi \leq \bigvee D \) implies that there is a \( \psi \in D \) such that \( \phi \leq \psi \).

The second property in this theorem says that elements of \( \Phi_f \) are “finite” in the sense of order or lattice theory. The theorem together with the density property above says then that \( \Phi \) is an algebraic lattice. This is sometimes also called a domain, although in domain theory the concept of domain can take slightly different meanings. But in any way this shows that information algebras are closely connected to domain theory, even if its focus is different from domain theory. It has been proved in [13,16] that the information algebra induced by a compact information system is compact. And its finite elements are just the closed sets \( C(X) \) with \( X \) finite.

The next theorem tells us that a compact information algebra can always be obtained by completion of an arbitrary information algebra.

**Theorem 3** [13]. Let \( \Phi \) be an information algebra, \( I(\Phi) \) its completion. Then \( (I(\Phi), \Phi) \) is a compact information algebra, whose finite elements are \( \Phi \).

Inversely, to a compact information algebra a compact information system can be associated as follows: The set of tokens is \( \Phi_f \). The consequence operator \( C_\Phi \) is defined for \( X \subseteq \Phi_f \) by

\[
C_\Phi(X) = \{ \psi \in \Phi_f : \psi \leq \bigvee X \}.
\]

This is indeed a compact consequence operator on \( \Phi_f \) [13,16]. Let us define \( A_\Phi = \{ \psi \in \Phi_f : \psi \leq \phi \} \). Then we have \( A_{\uparrow X} := C_\Phi(X) \). The associated entailment relation is simply \( X \vdash_\Phi \psi \) if, and only if, \( \psi \leq \bigvee X \). \((\Phi_f, \vdash_\Phi)\) is the compact information system associated with the compact information algebra \((\Phi, \Phi_f)\).

Let \( \phi = \bigvee X \) and \( \psi = \bigvee Y \). Then, in the information algebra induced by \((\Phi_f, \vdash_\Phi)\) combination is defined by

\[
C_\Phi(X) \otimes C_\Phi(Y) = C_\Phi(A_\Phi \cup A_\Phi) = A_{\phi \otimes \psi},
\]
since $\bigvee (A_\phi \cup A_\psi) = \phi \otimes \psi$. So, the mapping $\phi \mapsto A_\phi$ from $\Phi$ to $C(\Phi_f)$ is one-to-one and maintains combination. Also $e \mapsto A_e = C_\phi (\emptyset)$, since $\bigvee \emptyset = e$, and $z \mapsto A_z = \Phi_f$. Further, $\phi \in \Phi_f$ maps to $A_\phi = C_\phi (\{\phi\})$, i.e., to a finitely axiomatizable set. This shows that the information algebra $C(\Phi_f)$ induced by the information system $(\Phi_f, \vdash \Phi)$ is isomorph to the original finitary information algebra $(\Phi, \Phi_f)$. Thus, we have proved the following theorem:

**Theorem 4** [13]. If $(\Phi, \Phi_f)$ is a compact information algebra, then the system $(\Phi_f, \vdash \Phi)$ is a compact information system. The compact information algebra induced by this information system is isomorphic to $(\Phi, \Phi_f)$.

Here follow a few examples of compact information algebras.

1. **Subsets.** If intersection is the combination operation for subsets, then cofinite sets (sets of the form $S \setminus F$, with $F$ finite) are the finite elements of a compact information algebra. Dually, we may take union as combination. Then finite sets are the finite elements of the algebra.

2. **Binary strings.** If we add the infinite strings to $\{0, 1\}^*$ we get the set $\{0, 1\}^{**}$ of finite and infinite binary strings. We note that all strings (finite or not) may be identified with the family of their initial strings. Then $\{0, 1\}^{**}$ can be identified with the ideals of $\{0, 1\}^*$. Thus it becomes clear that $\{0, 1\}^{**}$ is the completion of $\{0, 1\}^*$, and that $\{0, 1\}^{**}$ is a compact information algebra with the finite strings as finite elements.

3. **Intervals.** We may consider intervals with rational bounds $-\infty \leq x \leq y \leq +\infty$ (we consider here $-\infty$ and $+\infty$ as rational numbers). These intervals form an information algebra under intersection. The algebra of intervals with real bounds is then its completion, since each such interval may be identified with the ideal of the rational intervals in which it is contained.

4. **Vector spaces.** Consider a vector space $V$ and define for $A \subseteq V, v \in V$ the entailment relation $A \vdash v$ if $v$ is an element of the linear manifold spanned by some finite set $B \subseteq A$. Then this generates the compact information algebra of linear manifolds with intersection as combination.

4. The probability structure

4.1. **Simple probability structures**

We consider now uncertain information, i.e., assumption-based information again. Let $\Omega$ be a set of possible assumptions, scenarios or interpretations and $X(\omega)$ the information asserted, if scenario $\omega$ is assumed. Let then

$$\omega \mapsto C(X(\omega))$$

be the corresponding mapping from $\Omega$ into the information algebra $L_C$. To each scenario $\omega$, the information $C(X(\omega))$ is associated. In order to study such mappings, we consider more generally an information algebra $\Phi$ and mappings $X: \Omega \rightarrow \Phi$. Similarly to (2.2), for
a hypothesis $\phi \in \Phi$, we may define the set of supporting scenarios
$$s_X(\phi) = \{\omega \in \Omega: \phi \leq X(\omega)\}.$$  

Note that we do at this place not eliminate inconsistent assumptions (see Section 2). This very important issue will be taken up in Section 4.5. For the time being we work for technical reasons in fact with quasi-supports and content ourselves to state that this will be the basis for the semantically more important supports.

If we have an assumption-based information structure $X: \Omega \rightarrow \Phi$, we may judge hypotheses $\phi \in \Phi$ by the support $s_X(\phi)$ allocated to them. The larger this support, the more credible or likely is a hypothesis. This is however a purely qualitative point of view. Not all scenarios $\omega \in \Omega$ may be equally likely. A hypothesis which is supported by many unlikely scenarios may finally be less credible than another one, which is supported by few scenarios only, but highly likely ones. In order to take this into consideration we may assign probabilities to scenarios. This is the subject of this section. If we assume $\Omega$ to be a finite or countably infinite set, then the mappings introduced above are fine. But already in applications like linear equations with Gaussian disturbances, this simple case is no more present. Then questions of measurability arise. And the general mappings considered above are too general to be useful. We need to restrict them to more reasonable categories. In this section we introduce first a very simple class of mappings, which will be generalized in the next section.

We assume a probability space $(\Omega, A, P)$ with a $\sigma$-algebra $A$ in $\Omega$ and a probability measure $P$ on it. Based on such a probability space we introduce first a finite probability structure.

**Definition 1.** A simple probability structure (s.p.s.) is given by
$$\left\{ \psi_1, \ldots, \psi_m \right\}$$
where the $\{A_1, \ldots, A_m\}$ form a decomposition of $\Omega$, $A_i \in A$, $P(A_i) > 0$ for $i = 1, \ldots, m$ and $\psi_i \in \Phi$ for $i = 1, \ldots, m$ and $\psi_i \neq \psi_j$ for $i \neq j$. The elements $\psi_i$ are called focal elements.

To a s.p.s. we associate a simple random variable (s.r.v.) $X : \Omega \rightarrow \Phi$, defined by
$$X(\omega) = \psi_i, \quad \forall \omega \in A_i.$$  
And we associate also a support mapping $s_X : \Phi \rightarrow A$ defined by
$$s_X(\phi) = \bigcup_{i: \psi_i \geq \phi} A_i = \{\omega: \phi \leq X(\omega)\}, \quad \forall \phi \in \Phi.$$  

We consider two s.r.v.s $X$ and $X'$ to be equal, $X = X'$, if $X(\omega) = X'(\omega)$ almost everywhere (a.e., i.e., except of a set of measure zero). Then, correspondingly, we take two support mappings $s$ and $s'$ to be equal $s = s'$, if for all $\phi \in \Phi$ the symmetric difference $s(\phi) \triangle s'(\phi)$ is a set of measure zero. So, to each s.r.v. $X$, a support mapping $s_X$ is associated and $X = X'$ implies $s_X = s_{X'}$. We write $H \subseteq a\, H' \ (H$ is almost surely contained in $H')$ if $H' \setminus H$ is of measure zero.
We denote by $X_s$ the set of all s.r.v.s (equal variables assumed to be identical) and by $S_s$ the corresponding support functions. Note that we have a one to one mapping $X \mapsto sX$ between the two sets. We are going to introduce an operation of combination into the two sets. We do this by defining a combination operation between two s.p.s.:

$$\{\psi_{11}, \ldots, \psi_{1r}\} \otimes \{\psi_{21}, \ldots, \psi_{2s}\} = \{\psi_{1}, \ldots, \psi_{m}\}$$

where $\psi_k = \psi_{1i} \otimes \psi_{2j}$ for some $i$ and $j$, and $A_k = \bigcup\{A_{1i} \cap A_{2j} : \psi_{1i} \otimes \psi_{2j} = \psi_k\}$. $P(A_k) > 0$ for $k = 1, \ldots, m$. This is again a s.p.s. Let $X$ and $sX$ be the s.r.v. and the support mapping associated to this s.p.s., whereas $X_1$, $X_2$ and $sX_1$, $sX_2$ are the s.r.v.s and the support mappings associated with the two original s.p.s. Then we define $X_1 \otimes X_2 := X$ and $sX_1 \otimes sX_2 := sX$. Note that $sX_1 \otimes sX_2 = sX_1 \otimes sX_2$. The s.r.v. $e(\omega) = e$ for all $\omega$ and the corresponding support mapping $s_e(\phi) = \emptyset$ for all $\phi \neq e$ and $s_e(e) = \Omega$ are the neutral elements of this combination operation, whereas the s.r.v. $z(\omega) = z$ and the corresponding support function $s_z(\phi) = \emptyset$ are the null elements of the combination. Clearly the operation is commutative and associative. Therefore $X_s$ and $S_s$ are two isomorphic information algebras.

We may express combination of s.r.v.s and support mappings also more directly, as the following lemma shows.

**Lemma 2.** For all $X_1, X_2 \in X_s$ and for all $\omega \in \Omega$, 

$$(X_1 \otimes X_2)(\omega) = X_1(\omega) \otimes X_2(\omega).$$

For all $s_1, s_2 \in S_s$ and for all $\phi \in \Phi$, 

$$(s_1 \otimes s_2)(\phi) = \bigcup_{\phi \leq \phi_1 \otimes \phi_2} \{s_1(\phi_1) \cap s_2(\phi_2)\}.\quad (4.2)$$

**Proof.** We have 

$$(X_1 \otimes X_2)(\omega) = \psi_{1i} \otimes \psi_{2j} = X_1(\omega) \otimes X_2(\omega)$$

for $\omega \in A_{1i} \cap A_{2j}$. So (4.1) holds everywhere.

Further let $X_1$ and $X_2$ be the s.r.v.s associated with $s_1$ and $s_2$ such that $s_1 \otimes s_2 = sX_1 \otimes sX_2$. Then, we obtain for all $\phi$, using the first part of the lemma

$$s_{X_1 \otimes X_2}(\phi) = \{\omega: \phi \leq X_1(\omega) \otimes X_2(\omega)\}$$

$$= \{\omega: \phi \leq \phi_1 \otimes \phi_2, \ \phi_1 \leq X_1(\omega), \ \phi_2 \leq X_2(\omega)\}$$

$$= \bigcup_{\phi \leq \phi_1 \otimes \phi_2} \{\{\omega: \phi_1 \leq X_1(\omega)\} \cap \{\omega: \phi_2 \leq X_2(\omega)\}\}$$

$$= \bigcup_{\phi \leq \phi_1 \otimes \phi_2} \{s_1(\phi_1) \cap s_2(\phi_2)\}. \quad \square$$

In both algebras $X_s$ and $S_s$ we have a partial order. The next lemma shows how this order is characterized.
Lemma 3. We have \( X' \subseteq X \) if, and only if, \( X'(\omega) \subseteq X(\omega) \) a.e. and, for \( s, s' \in \mathcal{S} \), we have \( s' \subseteq s \) if, and only if, for all \( \phi \in \Phi \) we have \( s'(\phi) \subseteq_{ax} s(\phi) \).

Proof. \( X' \subseteq X \) means \( X' \otimes X = X \), which in turn implies \( X'(\omega) \otimes X(\omega) = X(\omega) \) for almost all \( \omega \), hence \( X'(\omega) \subseteq X(\omega) \). The converse follows in the same way.

Let further \( X \) and \( X' \) be the s.r.v.s associated with the support mappings \( s \) and \( s' \). Then \( s' \subseteq s \) implies \( s' \otimes s = sX' \otimes X \otimes s = X = \omega \). Since the mapping \( X \mapsto sX \) is an isomorphism, this implies \( X' \otimes X = X \) or \( X' \subseteq X \). Therefore, we have, by the first part of the lemma, for all \( \phi \in \Phi \)

\[
s_{X'}(\phi) = \{ \omega : \phi \subseteq X'(\omega) \} \subseteq_{ax} \{ \omega : \phi \subseteq X(\omega) \} = s_X(\phi).
\]

Conversely, assume that the inclusion above holds for all \( \phi \). Take \( \phi = X'(\omega) \). Then it can be seen that \( \phi = X'(\omega) \subseteq X(\omega) \) for almost all \( \omega \). But this implies according to the first part of the lemma that \( X' \subseteq X \), hence \( s_X' \subseteq s_X \). \( \square \)

In order to avoid subsequently tedious measurability considerations, we associate a probability algebra \((B, \mu)\) to the probability space \((\Omega, \mathcal{A}, P)\) in a classical way \([10,11]\): Take \( \mathcal{N} \) to be the \( \sigma \)-ideal of \( P \)-null sets and define \( B = \mathcal{A}/\mathcal{N} \) to be the quotient algebra of the equivalence classes of measurable sets modulo \( \mathcal{N} \) (two sets \( A \) and \( A' \) are equivalent modulo \( \mathcal{N} \), if the symmetric difference \( A \triangle A' \in \mathcal{N} \)). Then \( B \) is a complete Boolean algebra with least element \( \bot \) and top element \( \top \). Furthermore, the Boolean algebra satisfies the so-called countable chain condition, which means that for every subset \( E \subseteq B \) there is a countable subset \( D \) of \( E \) such that \( \bigwedge E = \bigvee D \). There is also countable subset \( D \) of \( E \) such that \( \bigvee E = \bigwedge D \). And the mapping \( \pi : A \rightarrow B \) defined by \( \pi(A) = [A] \) (\( [A] \) denotes the equivalence class of \( A \)) is a \( \sigma \)-homomorphism. On \( B \) we define a measure \( \mu \) by \( \mu(\pi(A)) = P(A) \). This measure is positive, i.e., \( \mu(b) = 0 \) implies \( b = \bot \). \( (B, \mu) \) is the probability algebra associated with \((\Omega, \mathcal{A}, P)\).

Definition 2. A basic probability assignment (b.p.a.) is given by

\[
\begin{pmatrix}
\psi_1 & \cdots & \psi_m \\
b_1 & \cdots & b_m
\end{pmatrix}
\]

where the \( \{b_1, \ldots, b_m\} \) form a decomposition of \( B \), with \( \mu(b_i) > 0 \) and \( \psi_i \in \Phi \) for \( i = 1, \ldots, m \) and \( \psi_i \neq \psi_j \) for \( i \neq j \). The elements \( \psi_i \) are still called focal elements.

To each s.p.s. a b.p.a. with the same focal sets \( \psi_i \) and \( b_i = \pi(A_i) \) is associated. Each b.p.a. induces a mapping \( h : \{b_1, \ldots, b_m\} \rightarrow \Phi \), defined by \( h(b_i) = \psi_i \) and a mapping \( \rho : \Phi \rightarrow B \) defined by

\[
\rho(\phi) = \bigvee_{\phi \subseteq \psi_i} b_i.
\]  

(4.3)

We call the mapping \( h \) still a simple random variable (s.r.v.). Equal random variables \( X \) and \( X' \) correspond to the same \( h \) and induce the same \( \rho \). The mapping \( \rho \) is called an allocation of probability (a.o.p.). The set of s.r.v.s \( h \) will be denoted by \( \mathcal{H}_s \) and the set of a.o.p.s by \( \mathcal{R}_s \). We clearly have one-to-one mappings between \( \mathcal{H}_s \) and \( \mathcal{R}_s \), as well as between \( \mathcal{H}_s \)
and $\lambda$ and $\mathcal{R}$ and $\mathcal{S}$. Just as we have defined an operation of combination between s.p.s. we define such an operation between b.p.a.s:

$$\begin{bmatrix}
\psi_{11} & \cdots & \psi_{1r} \\
b_{11} & \cdots & b_{1r}
\end{bmatrix} \otimes \begin{bmatrix}
\psi_{21} & \cdots & \psi_{2s} \\
b_{21} & \cdots & b_{2s}
\end{bmatrix} = \begin{bmatrix}
\psi_1 & \cdots & \psi_m \\
b_1 & \cdots & b_m
\end{bmatrix}$$

where $\psi_k = \psi_{1j} \otimes \psi_{2j}$ for some $i$ and $j$, and $b_k = \bigvee\{b_{1j} \land b_{2j} : \psi_{1j} \otimes \psi_{2j} = \psi_k\} \neq \bot$ for $k = 1, \ldots, m$. This is clearly again a b.p.a. Simple probability structures and basic probability assignments are simply two different ways to describe essentially (up to sets of measure zero) the same situation of uncertainty. The associated s.r.v.s $X$ and $h$, the support mapping $s$ and the allocation of probability $\rho$ are still other ways to describe the same situation. We shall therefore henceforth take $X$ as the identifying element of the situation and write $h_X, s_X$ and $\rho_X$ to denote the corresponding other elements referring to the same situation. Note that $\rho_X = \pi \circ s_X$ and that

$$h_{X_1 \otimes h_{X_2}} = h_{X_1 \otimes X_2}, \quad s_{X_1 \otimes s_{X_2}} = s_{X_1 \otimes X_2},$$

$$\rho_{X_1 \otimes \rho_{X_2}} = \rho_{X_1 \otimes X_2}.$$

$h_e$, $s_e$ and $\rho_e$ are the neutral elements of the corresponding combination operation and $h_z$, $s_z$ and $\rho_z$ are the null elements. Thus, the mappings $X \mapsto h_X$, $X \mapsto s_X$ and $X \mapsto \rho_X$ are isomorphisms.

We may define the combination in $\mathcal{R}$ in a more direct way, as we did it for $\mathcal{S}$ in Lemma 2 above:

**Lemma 4.** For all $\rho_1, \rho_2 \in \mathcal{R}$ and for all $\phi \in \Phi$, $$(\rho_1 \otimes \rho_2)(\phi) = \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho_1(\phi_1) \land \rho_2(\phi_2)\}.$$ For the proof, we refer to [13, Theorem 7.17].

There is a partial order both in $\mathcal{H}$ and $\mathcal{R}$. These orders can be characterized as shown in the following lemma (which parallels Lemma 3). We say that a partition

$b = \{b_1, \ldots, b_m\}$

of $B$ is finer than another one $b' = \{b'_1, \ldots, b'_m\}$, $b' \leq b$, if for every $i$ there is a $j$ such that $b_i \leq b'_j$.

**Lemma 5.** If $h$ and $h'$ are s.r.v.s, then $h' \leq h$ if, and only if, $b' \leq b$, and $h'(b'_j) \leq h(b_i)$, whenever $b_i \leq b'_j$. If $\rho$ and $\rho'$ are a.o.p.s, then $\rho' \leq \rho$ if, and only if, for all $\phi \in \Phi$, $\rho'(\phi) \leq \rho(\phi)$.

**Proof.** If $h' \leq h$, then $h' \otimes h = h$, such that $b' \leq b$ and $h'(b'_j) \otimes h(b_i) = h(b_i)$, i.e., $h'(b'_j) \leq h(b_i)$, if $b_i \leq b'_j$. Conversely, if $b' \leq b$ and $h'(b'_j) \leq h(b_i)$, whenever $b_i \leq b'_j$, then $h'(b'_j) \otimes h(b_i) = h(b_i)$ and indeed $h' \otimes h = h$, or $h' \leq h$.

Let $\rho' \leq \rho$ such that $\rho' \otimes \rho = \rho$ or, for all $\phi \in \Phi$, $(\rho' \otimes \rho)(\phi) = \rho(\phi)$. Then (see Lemma 4 above),

$$(\rho' \otimes \rho)(\phi) = \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho'(\phi_1) \land \rho(\phi_2)\} = \rho(\phi).$$
We have therefore $\rho'(\phi_1) \land \rho(\phi_2) \leq \rho(\phi)$, if $\phi \leq \phi_1 \otimes \phi_2$. Take then $\phi_1 = \phi$ and $\phi_2 = e$, to obtain $\rho'(\phi) \land \rho(e) \leq \rho(\phi)$. But $\rho(e) = \top$, such that indeed $\rho'(\phi) \leq \rho(\phi)$. Conversely, if $\rho'(\phi) \leq \rho(\phi)$ for all $\phi \in \Phi$, then we have

$$
(\rho' \otimes \rho)(\phi) = \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho'(\phi_1) \land \rho(\phi_2)\}
\leq \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho(\phi_1) \land \rho(\phi_2)\} = (\rho \otimes \rho)(\phi) = \rho(\phi).
$$

But we have always $\rho \leq \rho' \otimes \rho$, such that, by what we just proved, $\rho(\phi) \leq (\rho' \otimes \rho)(\phi)$ for all $\phi$. Hence we conclude that $(\rho' \otimes \rho)(\phi) = \rho(\phi)$ for all $\phi$, and thus $\rho = \rho' \otimes \rho$ or $\rho' \leq \rho$. $\blacksquare$

We can assign probabilities to the support of any $\phi \in \Phi$. If $X$ is a s.r.v. then we define $ds_X(\phi) = P(s_X(\phi)) = \mu(\rho_X(\phi))$. The larger this probability, the more credible is $\phi$ in the light of the information given by the s.r.v. $X$ or the more is $\phi$ supported by this information. Therefore, $ds_X(\phi)$ is called the degree of support of $\phi$. We shall come back to degrees of support more in detail in Section 4.4. But before, we need to extend the notion of a random variable.

4.2. Random variables

Consider the algebra $X_\ast$ of simple random variables in a information algebra $\Phi$. If $X \subseteq X_\ast$ is an ideal in $X_\ast$, then we may consider $X$ as an element of the completion $I(X_\ast)$ of $X_\ast$. We call any such ideal $X \subseteq X_\ast$ a random variable (r.v.) in the information algebra $\Phi$. We remark immediately that, for all $\omega \in \Omega$,

$$
X(\omega) = \{X'(\omega): X' \in X\}
$$

is an ideal in $\Phi$ and $X(\omega)$ is an element of the completion of $I(\Phi)$ of $\Phi$. So, random variables may be considered as mappings from $\Omega$ into $I(\Phi)$. As before, we consider two r.v.s $X$ and $X'$ as equal, if in $I(\Phi)$ we have $X(\omega) = X'(\omega)$ for almost all $\omega$. Let $\mathcal{X}$ be the set of all random variables (equal variables identified), then $(\mathcal{X}, \mathcal{X})$ is a compact information algebra.

Next we associate support mappings $s_X$ with random variables $X$: For all $X \in \mathcal{X}$, the set

$$
s_X = \{s_X': X' \in X\}
$$

is an ideal in $S$, hence an element of the completion $I(S)$ of $S$. Let $I(S) = S$. Then $(S, \mathcal{S})$ is again a compact information algebra. We associate with $s_X$ a mapping $\Phi \to \mathcal{P}(\Omega)$ from $\Phi$ into the power sets of $\Omega$:

$$
s_X(\phi) = \{\omega \in \Omega: \phi \in X(\omega)\} = \bigcup_{X' \in X} s_X'(\phi).
$$

(4.4)
We claim that the mapping \( X \mapsto s_X \) is one-to-one. In fact, assume that \( s_X = s_Y \). Then we have for all \( \phi \in \Phi \)
\[
\{ \omega \in \Omega : \phi \in X(\omega) \} = \{ \omega \in \Omega : \phi \in Y(\omega) \}.
\]
Fix an \( \omega \) and take \( \phi = X'(\omega) \in X(\omega) \). Then the identity above implies \( X'(\omega) \in Y(\omega) \), such that \( X(\omega) \subseteq Y(\omega) \). In the same way we see that \( Y(\omega) \subseteq X(\omega) \). Hence we have \( X(\omega) = Y(\omega) \) for all \( \omega \in \Omega \), thus \( X = Y \). The map maps \( X \) also onto \( S \). Therefore, if we introduce an operation of combination into \( S \) by
\[
s_X \otimes s_Y := s_{X \otimes Y}
\]
the map becomes a homomorphism, hence an isomorphism between the compact information algebras \((X, X)\) and \((S, S)\).

The combination operation in \( S \) can, as before, be expressed more explicitly (compare (4.2)).

\section*{Lemma 6.}
For all \( \phi \in \Phi \), \( X, Y \in \mathcal{X} \),
\[
(s_X \otimes s_Y)(\phi) = \bigcup_{\phi_1, \phi_2} \{ s_X(\phi_1) \cap s_Y(\phi_2) \}.
\]
\section*{Proof.}
We have, by definition and by (4.4),
\[
(s_X \otimes s_Y)(\phi) = s_{X \otimes Y}(\phi) = \bigcup_{Z \in X \otimes Y} s_Z(\phi).
\]
Consider a \( \omega \in s_Z(\phi) \) such that \( \phi \leq Z'(\omega) \). But \( Z' \in X \otimes Y \) means that there are \( X' \in X \) and \( Y' \in Y \) such that \( Z' \leq X' \otimes Y' \), hence \( Z'(\omega) \leq X'(\omega) \otimes Y'(\omega) \) a.e. (Lemmas 3 and 2). Therefore we see that \( \omega \in s_{X' \otimes Y'}(\phi) \). Since \( X' \otimes Y' \in X \otimes Y \) if \( X' \in X \) and \( Y' \in Y \), we conclude that (Lemma 2)
\[
(s_X \otimes s_Y)(\phi) = \bigcup_{X' \in X, Y' \in Y} s_{X' \otimes Y'}(\phi) = \bigcup_{X' \in X, Y' \in Y} \bigcup_{\phi \leq \phi_1 \otimes \phi_2} \{ s_X(\phi_1) \cap s_Y(\phi_2) \}.
\]

The combination operation is closely related to the partial order. We can express this order also in different ways.

\section*{Lemma 7.}
The following statements are all equivalent: For \( X, Y \in \mathcal{X} \),
\begin{enumerate}
\item \( X \leq Y \),
\item for almost all \( \omega \in \Omega \), \( X(\omega) \subseteq Y(\omega) \) in \( I(\Phi) \),
\end{enumerate}
(3) \( s_X \leq s_Y \) in \( \mathcal{S} \).
(4) for all \( \phi \in \Phi \), \( s_X(\phi) \subseteq_{a \alpha} s_Y(\phi) \).

**Proof.** The equivalence between (1) and (3) follows from the isomorphism between \( X \) and \( \mathcal{S} \).

To prove the equivalence between statements (1) and (2) assume first that \( X \leq Y \), or if we look at \( X \) and \( Y \) as ideals, \( X \subseteq Y \). This means that for almost all \( \omega \) we have \( X'(\omega) \subseteq Y(\omega) \). Hence for almost all \( \omega \) we have \( X(\omega) \subseteq Y(\omega) \). Hence for almost all \( \omega \) we have \( X'(\omega) \subseteq Y(\omega) \), which means \( X(\omega) \subseteq Y(\omega) \) or \( X(\omega) \subseteq Y(\omega) \). Conversely, assume (2). Then \( X' \in Y \) for all \( X' \in X \), hence \( X = Y \). Hence for almost all \( \omega \) we have \( X(\omega) \subseteq Y(\omega) \) or \( X(\omega) \subseteq Y(\omega) \). Conversely, assume that for all \( \omega \in \Omega \), \( s_X(\omega) \subseteq_{a \alpha} s_Y(\omega) \). Then, by Lemma 6

\[
(s_X \otimes s_Y)(\phi) = \bigcup_{\phi \in \Phi \otimes \Phi} \left\{ s_X(\phi_1) \cap s_Y(\phi_2) \right\} \subseteq_{a \alpha} \bigcup_{\phi \in \Phi \otimes \Phi} \left\{ s_Y(\phi_1) \cap s_Y(\phi_2) \right\} = s_Y(\phi).
\]

But \( s_X \otimes s_Y \geq s_Y \) implies by the first part of the proof that \( (s_X \otimes s_Y)(\phi) \subseteq s_Y(\phi) \). Therefore we conclude that \( (s_X \otimes s_Y)(\phi) = s_Y(\phi) \) for all \( \phi \in \Phi \), hence \( s_X \otimes s_Y = s_Y \) or \( s_X \leq s_Y \). \( \square \)

The following theorem gives important properties of support mappings.

**Theorem 5.** A support mapping \( s_X : \Phi \to \mathcal{P}(\Omega) \) has the following properties:

(1) \( s_X(e) = \Omega \),
(2) For all \( \phi, \psi \in \Phi \), we have

\[
s_X(\phi \otimes \psi) = s_X(\phi) \cap s_X(\psi).
\] (4.5)

**Proof.** (1) is evident. We prove (2) first for simple random variables \( X' \). In order to do this, we note that

\[
\left\{ \omega: \phi \otimes \psi \leq X'(\omega) \right\} = \left\{ \omega: \phi, \psi \leq X'(\omega) \right\} = \left\{ \omega: \phi \leq X'(\omega) \right\} \cap \left\{ \omega: \psi \leq X'(\omega) \right\}.
\]

This shows then that \( s_X(\phi \otimes \psi) = s_X(\phi) \cap s_X(\psi) \). In the general case, we have therefore

\[
s_X(\phi \otimes \psi) = \bigcup_{X' \in X} s_X(\phi) \cap s_X(\psi) = \bigcup_{X' \in X} \left( s_X(\phi) \cap s_X(\psi) \right) = s_X(\phi) \cap s_X(\psi).
\] \( \square \)
As in the case of simple random variables we would like to use the support mapping \( s_X \) of a random variable \( X \) to determine the distribution function of \( X \), i.e., \( P(s_X(\phi)) \). This would measure the degree of support assigned to an information \( \phi \) by the random variable \( X \). Unfortunately however \( s_X(\phi) \) will not be measurable in general, and the probability \( P(s_X(\phi)) \) therefore not defined. We may nevertheless consider any measurable subset \( A \in \mathcal{A} \) of \( \Omega \), which is contained in \( s_X(\phi) \) as an “argument” supporting \( \phi \). Hence, the degree of support \( d s_X(\phi) \) for \( \phi \) should be at least \( P(A) \) for all such sets \( A \), i.e., \( d s_X(\phi) \geq P(A) \). In the absence of other information we may therefore tentatively define \( d s_X(\phi) \) as the supremum of all these probabilities,

\[
d s_X(\phi) = \sup \{ P(A): A \in \mathcal{A}, A \subseteq s_X(\phi) \}
\]

\( P_* \) is the inner probability measure of \( P \). We shall in the following subsection further study and justify this definition of the degree of support of elements of \( \phi \).

4.3. Allocation of probability

For simple random variables \( X' \) we have already defined the allocation of probability (a.o.p.) by \( \rho_{X'} = \pi \circ s_{X'} \). If we want to extend this definition to random variables \( X \), we have two possibilities:

(a) Either we first extend the map \( \pi: \mathcal{A} \rightarrow \mathcal{B} \) to the power set of \( \Omega \). This can, guided by (4.6), be done as follows: For all \( H \subseteq \Omega \) define,

\[
\rho_0(H) = \bigvee \{ \pi(A): A \in \mathcal{A}, A \subseteq H \}.
\]

We may then define \( \rho_X = \rho_0 \circ s_X \). Since, for all \( A \in \mathcal{A} \) we have \( \rho_0(A) = \pi(A) \), this is an extension of the definition of a.o.p. for s.r.v. to random variables (r.v.).

(b) Alternatively, we may consider for all \( X' \in X \) the probability \( \rho_{X'}(\phi) \) allocated to a fixed information \( \phi \). The probability \( \rho_X(\phi) \) allocated to \( \phi \) by \( X \) should be then at least \( \rho_{X'}(\phi) \) if \( X' \in X \). Then we may define \( \rho_X(\phi) \) as the supremum of all \( \rho_{X'}(\phi) \),

\[
\rho_X(\phi) = \bigvee_{X' \in X} \rho_{X'}(\phi).
\]

As the following theorem shows, the two definitions are in fact identical.

**Theorem 6.** For all \( \phi \in \Phi \),

\[
\rho_0(s_X(\phi)) = \bigvee_{X' \in X} \rho_{X'}(\phi).
\]

**Proof.** By definition, we have

\[
\rho_0(s_X(\phi)) = \bigvee \{ \pi(A): A \in \mathcal{A}, A \subseteq s_X(\phi) \}
\]

and \( \rho_{X'}(\phi) = \pi(s_{X'}(\phi)) \) for \( X' \in X \). By Lemma 7, \( X' \in X \), which means that \( X' \subseteq X \) in \( X' \), implies that \( s_{X'}(\phi) \subseteq s_X(\phi) \) for all \( \phi \). Since \( \rho_0 \) is monotone we conclude that
\[ \rho_X(\phi) = \rho_0(s_X(\phi)) \leq \rho_0(s_X(\phi)), \text{ hence} \]
\[ \rho_0(s_X(\phi)) \geq \bigvee_{X' \in X} \rho_{X'}(\phi) \quad (4.10) \]
for all \( \phi \in \Phi \).

Fix now \( \phi \in \Phi \). Then there exists a countable family of s.r.v.s \( \{X'_i \subseteq X, \ i = 1, 2, \ldots\} \) such that
\[ \bigvee_{i=1}^{\infty} \rho_{X'_i}(\phi) = \bigvee_{X' \in X} \rho_{X'}(\phi). \]
Let \( A_i = s_{X'_i}(\phi) \in \mathcal{A} \) and \( A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \). Then we have
\[ \bigvee_{i=1}^{\infty} \rho_{X'_i}(\phi) = \bigvee_{i=1}^{\infty} \pi(A_i) = \pi \left( \bigcup_{i=1}^{\infty} A_i \right) = \pi(A), \]
since \( \pi \) is a \( \sigma \)-homomorphism. Thus we have \( \bigvee_{X' \in X} \rho_{X'}(\phi) = \pi(A) \).

Further there exists a countable family \( \{B_i: B_i \in \mathcal{A}, B_i \subseteq s_X(\phi), \ i = 1, 2, \ldots\} \) such that
\[ \rho_0(s_X(\phi)) = \bigvee_{i=1}^{\infty} \pi(B_i) = \pi \left( \bigcup_{i=1}^{\infty} B_i \right) = \pi(B), \]
where \( B = \bigcup_{i=1}^{\infty} B_i \subseteq s_X(\phi) \) and \( B \in \mathcal{A} \).

We claim that \( A \Delta B \in \mathcal{N} \), such that \( \pi(A) = \pi(B) \). This would then prove the theorem.

We have already seen that \( \pi(A) \leq \pi(B) \) which implies that \( \pi(A \cup B) = \pi(B) \). Therefore we conclude that \( A \setminus B \in \mathcal{N} \). We can neglect this part and suppose that \( A \subseteq B \). Assume now that \( A \Delta B = B \setminus A \notin \mathcal{N} \). This means that \( P(B \setminus A) > 0 \). We define a s.r.v.
\[ X'_0(\omega) = \begin{cases} \phi & \text{for } \omega \in A_0 = B \setminus A, \\ e & \text{for } \omega \in \Omega \setminus A_0. \end{cases} \]
Then, for all \( \omega \in B \) we have \( \phi \leq X(\omega) \). Hence, for all \( \omega \in \Omega \) we have \( X'_0(\omega) \leq X(\omega) \), hence \( X'_0 \subseteq X \) or \( X'_0 \in X \) (Lemma 7). This implies
\[ \pi(A) = \bigvee_{X' \in X} \rho_{X'}(\phi) = \rho_{X'_0}(\phi) \vee \left( \bigvee_{X' \in X} \rho_{X'}(\phi) \right) = \pi(A_0) \vee \pi(A) = \pi(A_0 \cup A). \]
But this is a contradiction, since \( A_0 \cap A = \emptyset \) and \( P(A_0) > 0 \).

Therefore, for all \( \phi \in \Phi \), we have \( \pi(A) = \pi(B) \) which proves the theorem. \( \square \)

This theorem is a strong justification for the definition of the a.o.p. associated to a r.v. \( X \) by either of the two definitions given above. This justification is enforced by the following two results, which have been proved in [13].
Theorem 7 [13]. For all r.v.s $X, Y \in \mathcal{X}$, and for all $\phi \in \Phi$

$$\rho_{X \otimes Y}(\phi) = \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho_X(\phi_1) \land \rho_Y(\phi_2)\}. \quad (4.11)$$

As for support mappings, we may define an operation of combination between a.o.p.s by

$$\rho_X \otimes \rho_Y := \rho_{X \otimes Y}.$$

Theorem 7 allows then to express this operation in a more explicit form. It shows also that the algebra $\mathcal{R}_s$ of a.o.p.s of simple random variables is a subalgebra of the algebra of $\mathcal{R}$ of a.o.p.s of random variables. And the mapping $X \mapsto \rho_X$ is a homomorphism. The following lemma extends Lemma 7.

Lemma 8. The following statements are all equivalent:

1. $X \leq Y$ in $\mathcal{X}$,
2. $\rho_X \leq \rho_Y$ in $\mathcal{R}$,
3. for all $\phi \in \Phi$, $\rho_X(\phi) \leq \rho_Y(\phi)$.

Proof. (1) $\Rightarrow$ (2) follows since $X \leq Y$ means $X \otimes Y = Y$, hence $\rho_X \otimes \rho_Y = \rho_{X \otimes Y} = \rho_Y$, hence $\rho_X \leq \rho_Y$ in $\mathcal{R}$.

Assume then (2), that is $\rho_{X \otimes Y} = \rho_Y$. By Theorem 7 for all $\phi \in \Phi$,

$$\rho_{X \otimes Y}(\phi) = \bigvee_{\phi \leq \phi_1 \otimes \phi_2} \{\rho_X(\phi_1) \land \rho_Y(\phi_2)\} = \rho_Y(\phi).$$

If we take $\phi_1 = \phi$ and $\phi_2 = e$, then this implies that $\rho_X(\phi) \leq \rho_Y(\phi)$, hence (3).

Assume finally (3). This means $\rho_0(s_X(\phi)) \leq \rho_0(s_Y(\phi))$ for all $\phi$. This implies $s_X(\phi) \subseteq_{\text{a.s.}} s_Y(\phi)$. From Lemma 7 follows then that $X \leq Y$. \(\square\)

This lemma implies that the map $X \mapsto \rho_X$ from $\mathcal{X}$ to $\mathcal{R}$ is one-to-one and onto, hence an isomorphism.

Theorem 8 [13]. For all $X \in \mathcal{X}$ the corresponding a.o.p. $\rho_X$ has the following properties:

1. $\rho_X(e) = \top$.
2. For all $\phi, \psi \in \Phi$,

$$\rho_X(\phi \otimes \psi) = \rho_X(\phi) \land \rho_X(\psi). \quad (4.12)$$

As a consequence of (2) of this theorem we note that $\phi \leq \psi$ implies $\rho_X(\psi) = \rho_X(\phi \otimes \psi) = \rho(\phi) \land \rho(\psi)$, hence $\rho_X(\psi) \leq \rho_X(\phi)$.

Shafer [24,26] calls a map from a Boolean algebra into a probability algebra with the properties of Theorem 8 an allocation of probability. We adapt this terminology for the more general case of information algebras. Shafer shows that a.o.p.s are related to belief
functions, a term coined by him for the lower probabilities introduced by Dempster [7]. This connection will be pursued in the next subsection within our framework of r.v.s in information algebras.

4.4. Belief functions

Given a random variable $X$, we may define the degrees of support assigned by $X$ to an element $\phi \in \Phi$ by

$$ds_X(\phi) = \mu(\rho_X(\phi)).$$

In this way every $\phi \in \Phi$ has a degree of support assigned, independently of the measurability of $s_X(\phi)$. If we consider the degree of support $ds_X(\phi)$ which a r.v. $X$ assigns to an information $\phi$ as a mapping $ds_X : \Phi \to [0, 1]$, i.e., as a function of $\phi$, then we call $ds_X$ the support function associated with the r.v. $X$. It plays a role similar to a distribution function of a r.v. $X$. In this section we study these functions.

First we consider the case of a simple random variable $X'$ associated with a basic probability assignment (b.p.a.)

$$\{\psi_1 \ldots \psi_m \mid b_1 \ldots b_m\}.$$ 

To the focal elements $\psi_i$, we associate the probabilities $m_i = \mu(b_i)$. Then we have

$$m_i > 0 \text{ for } i = 1, \ldots, m, \quad \sum_{i=1}^m m_i = 1.$$ 

We obtain then the following support function, if we use (4.3):

$$s_{X'}(\phi) = \mu(\rho_X(\phi)) = \mu\left(\bigvee_{\phi \leq \psi_i} b_i\right) = \sum_{\phi \leq \psi_i} m_i.$$

This corresponds to the usual definition of belief functions on discrete frames in the Dempster–Shafer theory of evidence [25]. So, at least in the case of s.r.v. our support functions are belief functions.

In the case of a r.v. $X$ we have no more a b.p.a. to express the support function. But the following theorem shows characteristic properties of a support function $ds_X$ associated with a r.v. $X$.

**Theorem 9.** For all random variables $X \in \mathcal{X}$, the associated support function $s_X$ has the following properties:

1. $s_X(e) = 1$.
2. For any finite set $\phi_1, \ldots, \phi_m \ni \phi$, $m \geq 1$

$$ds_X(\phi) \geq \sum_{\phi \neq \phi_1 \leq \{1, \ldots, m\}} (-1)^{|I|+1} ds_X(\bigotimes_{i \in I} \phi_i).$$

(4.14)
Proof. (1) follows immediately from (1) of Theorem 8.
In order to prove (2) we note first that $\rho_X(\phi_i) \leq \rho_X(\phi)$ for all $i = 1, \ldots, m$. Hence

$$\rho_X(\phi_1) \lor \cdots \lor \rho_X(\phi_m) \leq \rho_X(\phi).$$

From this we obtain

$$ds_X(\phi) = \mu(\rho_X(\phi) \lor \cdots \lor \rho_X(\phi_m)).$$

By the well-known inclusion-exclusion formula from probability theory the right-hand side of this inequality equals

$$\mu(\rho_X(\phi_1) \lor \cdots \lor \rho_X(\phi_m)) = \sum_{\emptyset \neq I \subseteq \{1, \ldots, m\}} (-1)^{|I|+1} \mu(\bigwedge_{i \in I} \rho_X(\phi_i)).$$

Here we have used (2) of Theorem 8. If we note that $\mu(\bigotimes_{i \in I} \rho_X(\phi_i)) = ds_X(\bigotimes_{i \in I} \phi_i)$ Eq. (4.14) follows.

A function satisfying (4.14) is called monotone to the order $\infty$. This, together with (1) of Theorem 9 is a characterizing property of belief functions, as has been shown in [24, 26]. In fact Shafer showed that any function having these properties is the support function associated to an allocation of probability. Again Shafer considered only Boolean structures $\Phi$. So, support functions on information algebras $\Phi$ are genuine generalizations of belief functions. It remains to show that as with Boolean structures $\Phi$ any function having the properties of Theorem 9 with respect to an information algebra $\Phi$ is induced by some allocation of probability on $\Phi$, hence some random variable in $\Phi$.

4.5. Normalization

So far we have neglected an important issue: A r.v. $X$ may have the value $z$ for some assumptions $\omega$, $X(\omega) = z$. Assumptions, which map to $z$ are however to be considered as inconsistent with the information expressed by the variable $X$. Let

$$I_X = \{ \omega \in \Omega : X(\omega) = z \}$$

be the set of inconsistent assumption, scenarios or interpretations (see Section 2). These scenarios have to be eliminated and the probability has to be conditioned on the consistent assumptions $C_X = \Omega \setminus I_X$. To an arbitrary r.v. $X$ we associate therefore its normalized version $X^*$ defined as the restriction of $X$ to $C_X$. But at the same time the original probability space $(\Omega, A, P)$ is replaced by the new probability space $(C_X, A \cap C_X, P_X)$, where $A \cap C_X$ is the $\sigma$-algebra of sets $A \cap C_X$, for $A \in A$ and $P_X$ is defined as

$$P_X(A \cap C_X) = \frac{P^*(A \cap C_X)}{P^*(C_X)}.$$
where $P^*$ is the outer probability measure [20]. We suppose here that $P^*(C_X) > 0$. Note that for simple random variables $C_X$ is measurable, so that the outer probability equals the probability.

A r.v. with $X(\omega) \neq z$ for all $\omega$ is called normalized. Now, even if we start with normalized r.v.s $X_1$ and $X_2$, their combination may be no more normalized. The combination can introduce inconsistencies, which have to be eliminated. Therefore, we define a new operation of combination,

$$X_1 \otimes X_2 = (X_1 \otimes X_2)^\downarrow.$$

This is called normalized combination. We may consider the set $\mathcal{H}^\downarrow$ of normalized r.v.s on a probability space $(\Omega, A, P)$ and verify that this set forms an information algebra [13].

We turn now to allocations of probabilities. For a normalized r.v. $X$ we have $\rho_X(z) = \perp$. An a.o.p. satisfying this condition is called a normalized allocation. We may normalize an arbitrary a.o.p. $\rho$ by

$$\rho^\downarrow(\phi) = \rho(\phi) \wedge \rho^\circ(\phi),$$

where $\wedge$ denotes complementation in $B$. $\rho^\downarrow$ is to be considered as a mapping from $\Phi$ into the probability algebra $(\mathcal{B} \wedge \rho^\circ(z), \mu')$, where $\mathcal{B} \wedge \rho^\circ(z)$ is the Boolean algebra of elements $b \wedge \rho^\circ(z)$, for all $b \in \mathcal{B}$ and $\mu'$ is the measure defined by

$$\mu'(b \wedge \rho^\circ(z)) = \frac{\mu(b \wedge \rho^\circ(z))}{\mu(\rho^\circ(z))}.$$

As for r.v.s we may define the corresponding normalized combination for normalized a.o.p.s

$$\rho_1 \otimes \rho_2 = (\rho_1 \otimes \rho_2)^\downarrow.$$

The set of normalized a.o.p.s $\mathcal{R}^\downarrow$ forms then under this normalized combination again an information algebra [13].

For normalized r.v.s $X$ (or a.o.p.s) the corresponding support function $ds_X$ satisfies the condition

$$ds_X(z) = 0.$$

Such support functions are called normalized.

Normalized combination for independent simple random variables corresponds to Dempster’s rule [7,25]. Two simple probabilistic structures

$$\begin{bmatrix}
\psi_{11} & \ldots & \psi_{1r} \\
A_{11} & \ldots & A_{1r}
\end{bmatrix}, \quad \begin{bmatrix}
\psi_{21} & \ldots & \psi_{2s} \\
A_{21} & \ldots & A_{2s}
\end{bmatrix}$$

are called orthogonal, if $A_{1i} \cap A_{2j} \neq \emptyset$, and

$$P(A_{1i} \cap A_{2j}) = P(A_{1i}) \cdot P(A_{2j})$$

for all $i = 1, \ldots, s$ and $j = 1, \ldots, r$. Two simple random variables $X_1$ and $X_2$ are called independent, if they are associated to two orthogonal simple probability structures.
have then
\[ ds_{X_1 \oplus X_2}(\phi) = \sum_{\phi \leq \psi_k} m_k, \]
where \( \psi_k = \psi_{1i} \otimes \psi_{2j} \neq z \) for some \( i, j \),
\[ m_0 = \sum_{i,j: \psi_{1i} \otimes \psi_{2j} = z} m_{1i} \cdot m_{2j} \]
and
\[ m_k = \frac{1}{m_0} \sum_{i,j: \psi_{1i} \otimes \psi_{2j} = \psi_k} m_{1i} \cdot m_{2j}. \]

This is Dempster's rule extended to information algebras. It may also be generalized to general independent r.v.s.

5. Conclusion

In this paper information systems are proposed as general abstract structures to formulate information and to deduce consequences of information. This covers many formalisms for describing information, especially many kind of logics. But it covers as well numerical systems like linear equations, linear inequalities or more general algebraic systems. We have shown that introducing assumptions into information systems permits to specify uncertain information. Again this covers many well-known systems such as assumption-based propositional logic as used in ATMS (Assumption-based Truth Maintenance Systems) [6,9]. But also numerical systems with disturbances or measurement errors, as considered, for example, in statistics [19].

It turns out that information systems induce an interesting algebraic structure which we call an information algebra. This algebra describes the operation of combination or aggregation of pieces of information. It induces also a partial order of information, that is, a qualitative measure of information content. Information algebras as defined in this paper are semilattices. But we want to stress, that for simplicity's sake we neglected one important issue here: Information refers to specified questions. These questions are connected in some network or lattice. And information must be not only combined in this network, but also focused on the questions of particular interest. Here we took only information with respect to one single question into considerations. If a network of questions is modeled, then a more interesting algebraic structure arises [13].

Information systems represent the logic-part of argumentation. Most important is that assumption-based information systems or the related information algebras prove to be very natural structures to introduce probability. This is done by defining random variables with values in information algebras. This requires some thought, since information algebras are not Boolean structures. But it is remarkable that random variables in information algebras form themselves information algebras and that the theory of information algebras permits in a bootstrap procedure to define general random variables with values in information algebras.
This gives us a very satisfactory theory. If we turn to the distribution function of these random variables we find out, that they are monotone of order $\infty$. This replaces additivity of probability measures, which makes no sense in information algebras, because they are not Boolean. This property qualifies distribution function, which here are called support functions, as a genuine generalization of belief functions in the sense of Dempster–Shafer (DS) theory of evidence. Thus this theory finds here again a clear foundation in probability theory. Further, since Boolean algebras are information algebras too, DS-theory, as presented here, proves to be a genuine extension of classical probability theory. This is then the probabilistic part of argumentation.

To conclude let us stress that many open problems remains in this promising theory. A very important issue in view of applications is the question of the independence and conditional independence between random variables in information algebras. This in turn is related to structures which allow for so-called “local computation schemes” which are well-known for Bayesian networks, but which can be generalized to much more general systems as has been shown in [27] (see also [13,15]).

Acknowledgements

I thank the anonymous referee for his critical remarks and corrections. My thanks go also to B. Anrig for his comments and proof reading of this paper.

References