draft – 2012-05-21

# A Probabilistic Semantics for Abstract Argumentation

**Matthias Thimm**[1]

**Abstract.** Classical semantics for abstract argumentation frameworks are usually defined in terms of extensions or, more recently, labelings. That is, an argument is either regarded as accepted with respect to a labeling or not. In order to reason with a specific semantics one takes either a credulous or skeptical approach, i. e. an argument is ultimately accepted, if it is accepted in one or all labelings, respectively. In this paper, we propose a more general approach for a semantics that allows for a more fine-grained differentiation between those two extreme views on reasoning. In particular, we propose a probabilistic semantics for abstract argumentation that assigns probabilities or *degrees of belief* to individual arguments. We show that our semantics generalizes the classical notions of semantics and we point out interesting relationships between concepts from argumentation and probabilistic reasoning. We illustrate the usefulness of our semantics on an example from the medical domain.

## 1 Introduction

The field of computational models of argumentation [18] is concerned with non-monotonic reasoning mechanisms that focus on the role of arguments. An argument is an entity that represents some grounds to believe in a certain statement and that can be in conflict with arguments establishing contradictory claims. The most commonly used framework to talk about general issues of argumentation is that of abstract argumentation [5]. In abstract argumentation, arguments are represented as atomic entities and the interrelationships between different arguments are modeled using an attack relation. Abstract argumentation has been thoroughly investigated in the past fifteen years and there is quite a lot of work on, e. g. extending abstract argumentation frameworks [8, 12, 7] and, in particular, semantical issues [3, 4, 2, 19]. Several different kinds of semantics for abstract argumentation frameworks have been proposed that highlight different aspects of argumentation. Usually, semantics are given to abstract argumentation frameworks in terms of extensions or, more recently, labelings. For a specific labeling an argument is either accepted, not accepted, or undecided. In a fixed semantical context, there is usually a set of labelings that is consistent with the semantical context. In order to reason with a semantics one has to take either a credulous or skeptical perspective. That is, an argument is ultimately accepted wrt. a semantics if the argument is accepted by at least one labeling consistent with that semantics (the credulous perspective) or if the argument is accepted by all labelings consistent with the semantics (the skeptical perspective). This extreme points of views may result in undesired results as in extreme cases the set of credulously accepted arguments may contain nearly the whole set of arguments and the set of skeptically accepted set of arguments may be nearly empty.

In this paper we propose a new way to assign semantics to abstract argumentation frameworks. More precisely, instead of using labelings we use probability functions on subsets of arguments as interpretations and define a probabilistic satisfaction relation that generalizes the notion of a complete labeling. In contrast to other works that combine abstract argumentation with quantitative uncertainty [12, 8, 6, 7, 11, 10, 1] we do not extend the underlying notion of an abstract argumentation framework but assess its inherent uncertainty using a more general semantics. In order to reason with this semantics we adopt notions from probabilistic reasoning for reasoning with sets of probability functions. We show that probabilistic semantics allow for a more fine-grained view on the relationships of arguments within an abstract argumentation framework.

On a more wider perspective, this paper also gives some first insights on the relationships between two of the most important subfields of artificial intelligence, namely argumentation and probabilistic reasoning [16, 15]. In particular, we show that the grounded labeling in abstract argumentation corresponds to the maximum entropy model in probabilistic reasoning (wrt. our probabilistic semantics for abstract argumentation frameworks).

The rest of this paper is organized as follows. In Section 2 we give a brief overview on abstract argumentation and exemplify the problems raised above to motivate our approach. In Section 3 we introduce a probabilistic semantics for abstract argumentation frameworks and discuss its properties. We continue in Section 4 with a comparison of the probabilistic semantics and classical semantics and show interesting relationships between notions from argumentation and probabilistic reasoning. In Section 5 we illustrate the usefulness of the approach with a short example and discuss related work in Section 6. We conclude in Section 7 with a summary and discussion.

## 2 Abstract Argumentation

*Abstract argumentation frameworks* [5] take a very simple view on argumentation as they do not presuppose any internal structure of an argument. Abstract argumentation frameworks only consider the interactions of arguments by means of an attack relation between arguments.

**Definition 1 (Abstract Argumentation Framework)** *An* abstract argumentation framework $\mathsf{AF}$ *is a tuple* $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ *where* $\mathsf{Arg}$ *is a set of arguments and* $\rightarrow$ *is a relation* $\rightarrow \subseteq \mathsf{Arg} \times \mathsf{Arg}$.

For two arguments $\mathcal{A}, \mathcal{B} \in \mathsf{Arg}$ the relation $\mathcal{A} \rightarrow \mathcal{B}$ means that argument $\mathcal{A}$ attacks argument $\mathcal{B}$. Abstract argumentation frameworks can be concisely represented by directed graphs, where arguments are represented as nodes and edges model the attack relation.

**Example 1** *Consider the abstract argumentation framework* $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ *depicted in Fig. 1. Here it is* $\mathsf{Arg} = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$ *and* $\rightarrow = \{(\mathcal{A}_1, \mathcal{A}_2), (\mathcal{A}_2, \mathcal{A}_1), (\mathcal{A}_2, \mathcal{A}_3), (\mathcal{A}_3, \mathcal{A}_4), (\mathcal{A}_4, \mathcal{A}_5), (\mathcal{A}_5, \mathcal{A}_4), (\mathcal{A}_5, \mathcal{A}_3)\}$.

[1] Institute for Web Science and Technologies Universität Koblenz-Landau, Germany

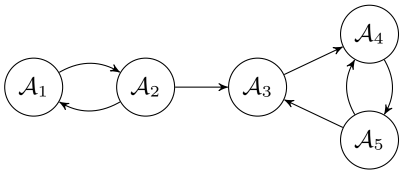


**Figure 1.** A simple argumentation framework

Semantics are usually given to abstract argumentation frameworks by means of extensions [5] or labelings [19]. An *extension* $E$ of an argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ is a set of arguments $E \subseteq \mathsf{Arg}$ that gives some coherent view on the argumentation underlying $\mathsf{AF}$. A labeling $L$ is a function $L : \mathsf{Arg} \rightarrow \{\mathsf{in}, \mathsf{out}, \mathsf{undec}\}$ that assigns to each argument $\mathcal{A} \in \mathsf{Arg}$ either the value $\mathsf{in}$, meaning that the argument is accepted, $\mathsf{out}$, meaning that the argument is not accepted, or $\mathsf{undec}$, meaning that the status of the argument is undecided. Let $\mathsf{in}(L) = \{\mathcal{A} \mid L(\mathcal{A}) = \mathsf{in}\}$ and $\mathsf{out}(L)$ resp. $\mathsf{undec}(L)$ be defined analogously. As extensions can be characterized by the arguments that labeled $\mathsf{in}$ in some labeling, we restrain our attention to labelings henceforth. In order to distinguish extension- and labeling-based semantics to the probabilistic semantics in the next section we denote the former *classical semantics*.

In the literature [5, 4] a wide variety of different types of classical semantics has been proposed. Arguably, the most important property of a semantics is its admissibility. A labeling $L$ is called *admissible* if and only if for all arguments $\mathcal{A} \in \mathsf{Arg}$

1. if $L(\mathcal{A}) = \mathsf{out}$ then there is $\mathcal{B} \in \mathsf{Arg}$ with $L(\mathcal{B}) = \mathsf{in}$ and $\mathcal{B} \rightarrow \mathcal{A}$, and
2. if $L(\mathcal{A}) = \mathsf{in}$ then $L(\mathcal{B}) = \mathsf{out}$ for all $\mathcal{B} \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$,

and it is called *complete* if, additionally, it satisfies

3. if $L(\mathcal{A}) = \mathsf{undec}$ then there is no $\mathcal{B} \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$ and $L(\mathcal{B}) = \mathsf{in}$ and there is a $\mathcal{B}' \in \mathsf{Arg}$ with $\mathcal{B}' \rightarrow \mathcal{A}$ and $L(\mathcal{B}') \neq \mathsf{out}$.

The intuition behind admissibility is that an argument can only be accepted if there are no attackers that are accepted and if an argument is not accepted then there has to be some reasonable grounds. The idea behind the completeness property is that the status of argument is only $\mathsf{undec}$ if it cannot be classified as $\mathsf{in}$ or $\mathsf{out}$. Different types of classical semantics can be phrased by imposing further constraints.

**Definition 2** *Let* $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ *be an abstract argumentation framework and* $L : \mathsf{Arg} \rightarrow \{\mathsf{in}, \mathsf{out}, \mathsf{undec}\}$ *a complete labeling.*

- $L$ *is* grounded *if and only if* $\mathsf{in}(L)$ *is minimal.*
- $L$ *is* preferred *if and only if* $\mathsf{in}(L)$ *is maximal.*
- $L$ *is* stable *if and only if* $\mathsf{undec}(L) = \emptyset$*.*
- $L$ *is* semi-stable *if and only if* $\mathsf{undec}(L)$ *is minimal.*

*All statements on minimality/maximality are meant to be with respect to set inclusion.*

Note that a grounded labeling is uniquely determined and always exists [5]. Besides the above mentioned types of classical semantics there are a lot of further proposals such as *CF2 semantics* [3]. However, in this paper we focus on complete, grounded, preferred, stable, and semi-stable semantics.

**Example 2** *We continue Ex. 1. Consider the labeling* $L$ *defined via*

$$L(\mathcal{A}_1) = \mathsf{in} \qquad L(\mathcal{A}_2) = \mathsf{out} \qquad L(\mathcal{A}_3) = \mathsf{out}$$
$$L(\mathcal{A}_4) = \mathsf{out} \qquad L(\mathcal{A}_5) = \mathsf{in}$$

*Clearly,* $L$ *is an admissible labeling as it satisfies properties 1.) and 2.) from above. Additionally, it is complete and also preferred, stable, and semi-stable. Furthermore, consider the labeling* $L'$ *defined via*

$$L'(\mathcal{A}_1) = \mathsf{out} \qquad L'(\mathcal{A}_2) = \mathsf{in} \qquad L'(\mathcal{A}_3) = \mathsf{out}$$
$$L'(\mathcal{A}_4) = \mathsf{in} \qquad L'(\mathcal{A}_5) = \mathsf{out}$$

*The labeling* $L'$ *is also admissible, complete, preferred, stable, and semi-stable. Note, that the grounded labeling* $L_g$ *is defined via* $L_g(\mathcal{A}_1) = L_g(\mathcal{A}_2) = L_g(\mathcal{A}_3) = L_g(\mathcal{A}_4) = L_g(\mathcal{A}_5) = \mathsf{undec}$*.*

As one can see in the above example, most semantics are *multi-extension* semantics. That is, there is not always a unique labeling induced by the semantics. In order to reason with multi-extension semantics, usually, one takes either a credulous or skeptical perspective. That is, an argument $\mathcal{A}$ is *credulously inferred* with semantics $\mathcal{S} \in \{\text{complete}, \text{preferred}, \text{stable}, \text{semi-stable}\}$ if there is a $\mathcal{S}$-labeling $L$ with $L(\mathcal{A}) = \mathsf{in}$. An argument $\mathcal{A}$ is *skeptically inferred* with semantics $\mathcal{S}$ if for all $\mathcal{S}$-labelings $L$ it holds that $L(\mathcal{A}) = \mathsf{in}$. Taking either a credulous or skeptical perspective is a crucial choice as the set of inferred arguments might change drastically.

**Example 3** *We continue Ex. 2. Besides* $L$ *and* $L'$ *there is also another labeling* $L''$ *that is admissible, complete, preferred, stable, and semi-stable:*

$$L''(\mathcal{A}_1) = \mathsf{out} \qquad L''(\mathcal{A}_2) = \mathsf{in} \qquad L''(\mathcal{A}_3) = \mathsf{out}$$
$$L''(\mathcal{A}_4) = \mathsf{out} \qquad L''(\mathcal{A}_5) = \mathsf{in}$$

*With respect to complete, preferred, stable, and semi-stable semantics, it follows that no argument is skeptically inferred and all arguments but* $\mathcal{A}_3$ *are credulously inferred.*

The example above shows that the difference of skeptical and credulous inference may be huge. Consequently, it is hard to assess the quality or strength of argument in an argumentation framework if only those types of inference are considered, cf. [13]. Imagine that the arguments in Ex. 1 are interpreted within a decision-support system in the medical domain. That is, the arguments $\mathcal{A}_1, \ldots, \mathcal{A}_5$ represent different drugs for a specific disease and an attack means a negative "influence" of one drug to another. In this system, a decision comprises a set of drugs that are used for treatment and the question is how to select this set? With credulous semantics the recommendation is to administer almost all drugs and with skeptical semantics the recommendation is to administer no drug. None of these recommendations seem appropriate in the example. In particular, administering no drug at all may not be possible as some action may be required to be performed. Another possible way to select the set of drugs is to select the drugs from one specific labeling. But then the question arises which labeling to chose?

## 3 Probabilistic Semantics

In order to get a more fine-grained view on the status of arguments we propose a new semantics that generalizes classical semantics and is based on a probabilistic interpretation of arguments. For that, we need some further notation. Let $2^{\mathcal{X}}$ denote the power set of a set $\mathcal{X}$.

**Definition 3** *Let* $\mathcal{X}$ *be some finite set. A* probability function $P$ *on* $\mathcal{X}$ *is a function* $P : 2^{\mathcal{X}} \rightarrow [0, 1]$ *that satisfies*

1. $P(\mathcal{X}) = 1$ *and*
2. $P(X_1 \cup X_2) = P(X_1) + P(X_2)$ *for* $X_1, X_2 \subseteq \mathcal{X}$*,* $X_1 \cap X_2 = \emptyset$*.*

For $x \in \mathcal{X}$ we write $P(x)$ instead of $P(\{x\})$. Here, a probability function is a function on the set of subsets of some (finite) set with two characteristic properties. First, the function must be *normalized*, i. e., the whole set must have probability one (property 1 above). Second, the probability of the union of two disjoint set is the sum of the probabilities of each set (property 2 above). These two properties are also called the *Kolmogorov properties of probability* [9]. The following observation is easy to see and the proof can be found e. g. in [15].

**Proposition 1** *For $X \subseteq \mathcal{X}$ and a probability function $P$ on $\mathcal{X}$ it holds*

$$P(X) = \sum_{x \in X} P(x)$$

Due to the above proposition a probability function can be defined just by defining the probabilities for each $x \in \mathcal{X}$.

A probability function is usually used to model statistical events. Then $\mathcal{X}$ is the set of all possible atomic events and subset $X$ of $\mathcal{X}$ represents the disjunction of the events in $X$. Given that $\mathcal{X}$ contains all possible events, property 1 above says that one event has to occur and property 2 states that atomic events are mutually exclusive.

In this paper, we use another interpretation for probability, that of *subjective probability* [15]. There, a probability $P(X)$ for some $X \subseteq \mathcal{X}$ denotes the *degree of belief* we put into $X$. Then a probability function $P$ can be seen as an epistemic state of some agent that has uncertain beliefs with respect to $\mathcal{X}$. In probabilistic reasoning [16, 15], this interpretation of probability is widely used to model uncertain knowledge representation and reasoning.

In the following, we consider probability functions on sets of arguments of an abstract argumentation frameworks. Let $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ be some fixed abstract argumentation framework and let $\mathcal{E} = 2^{\mathsf{Arg}}$ be the set of all sets of arguments. Let now $\mathcal{P}_{\mathsf{AF}}$ be the set of probability functions of the form $P : 2^{\mathcal{E}} \rightarrow [0, 1]$. A probability function $P \in \mathcal{P}_{\mathsf{AF}}$ assigns to each set of possible extensions of $\mathsf{AF}$ a probability, i. e., $P(e)$ for $e \in \mathcal{E}$ is the probability that $e$ is an extension and $P(E)$ for $E \subseteq \mathcal{E}$ is the probability that any of the sets in $E$ is an extension. In particular, note the difference between e. g. $P(\{\mathcal{A}, \mathcal{B}\}) = P(\{\{\mathcal{A}, \mathcal{B}\}\})$ and $P(\{\{\mathcal{A}\}, \{\mathcal{B}\}\})$ for arguments $\mathcal{A}, \mathcal{B}$. While the former denotes the probability that $\{\mathcal{A}, \mathcal{B}\}$ is an extension the latter denotes the probability that $\{\mathcal{A}\}$ or $\{\mathcal{B}\}$ is an extension. In general, it holds $P(\{\mathcal{A}, \mathcal{B}\}) \neq P(\{\{\mathcal{A}\}, \{\mathcal{B}\}\})$.

For $P \in \mathcal{P}_{\mathsf{AF}}$ and $\mathcal{A} \in \mathsf{Arg}$ we abbreviate

$$P(\mathcal{A}) = \sum_{\mathcal{A} \in e \subseteq \mathsf{Arg}} P(e) \quad .$$

Given some probability function $P$, the probability $P(\mathcal{A})$ represents the degree of belief that $\mathcal{A}$ is in an extension (according to $P$), i. e., $P(\mathcal{A})$ is the sum of the probabilities of all possible extensions that contain $\mathcal{A}$. The set $\mathcal{P}_{\mathsf{AF}}$ contains all possible views one can take on the arguments of an abstract argumentation framework $\mathsf{AF}$.

**Example 4** *We continue Ex. 1. Consider the function $P \in \mathcal{P}_{\mathsf{AF}}$ defined via $P(\{\mathcal{A}_1, \mathcal{A}_3, \mathcal{A}_5\}) = 0.3$, $P(\{\mathcal{A}_1, \mathcal{A}_4\}) = 0.45$, $P(\{\mathcal{A}_5, \mathcal{A}_2\}) = 0.1$, $P(\{\mathcal{A}_2, \mathcal{A}_4\}) = 0.15$, and $P(e) = 0$ for all remaining $e \in \mathcal{E}$. Due to Prop. 1 the function $P$ is well-defined as e. g.*

$$\begin{aligned} &P(\{\{\mathcal{A}_5, \mathcal{A}_2\}, \{\mathcal{A}_2, \mathcal{A}_4\}, \{\mathcal{A}_3\}\}) \\ &\quad = P(\{\mathcal{A}_5, \mathcal{A}_2\}) + P(\{\mathcal{A}_2, \mathcal{A}_4\}) + P(\{\mathcal{A}_3\}) \\ &\quad = 0.1 + 0.15 + 0 = 0.25 \end{aligned}$$

*Therefore, $P$ is a probability function according to Def. 3. According to $P$ the probabilities of each argument of $\mathsf{AF}$ compute to $P(\mathcal{A}_1) = 0.75$, $P(\mathcal{A}_2) = 0.25$, $P(\mathcal{A}_3) = 0.3$, $P(\mathcal{A}_4) = 0.6$, and $P(\mathcal{A}_5) = 0.4$.*

In the following, we are only interested in those probability functions of $\mathcal{P}_{\mathsf{AF}}$ that agree with our intuition on the interrelationships of arguments and attack. For example, if an argument $\mathcal{A}$ is not attacked we should completely believe in its validity if no further information is available. We propose the following notion of *justifiability* to describe this intuition.

**Definition 4** *A probability function $P \in \mathcal{P}_{\mathsf{AF}}$ is called $p$-justfiable wrt. $\mathsf{AF}$, denoted by $P \models_J \mathsf{AF}$, if it satisfies for all $\mathcal{A} \in \mathsf{Arg}$*

1. *$P(\mathcal{A}) \leq 1 - P(\mathcal{B})$ for all $\mathcal{B}, \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$ and*
2. *$P(\mathcal{A}) \geq 1 - \sum_{\mathcal{B} \in \mathcal{F}} P(\mathcal{B})$ where $\mathcal{F} = \{\mathcal{B} \mid \mathcal{B} \rightarrow \mathcal{A}\}$.*

*Let $\mathcal{P}^J_{\mathsf{AF}}$ be the set of all p-justifiable probability functions wrt. $\mathsf{AF}$.*

The notion of $p$-justifiability generalizes the concept of complete semantics to the probabilistic setting. Property 1.) says that the degree of belief we assign to an argument $\mathcal{A}$ is bounded from above by the inverse degrees of belief we put into the attackers of $\mathcal{A}$. As a special case, note that if we completely believe in an attacker of $\mathcal{A}$, i. e. $P(\mathcal{B}) = 1$ for some $\mathcal{B}$ with $\mathcal{B} \rightarrow \mathcal{A}$, then it follows $P(\mathcal{A}) = 0$. This corresponds to property 1.) of a complete labeling, see Section 2. Property 2.) of Def. 4 says that the degree of belief we assign to an argument $\mathcal{A}$ is bounded from below by the inverse of the sum of the degrees of belief we we put into the attacks of $\mathcal{A}$. As a special case, note that if we completely disbelieve in all attackers of $\mathcal{A}$, i. e. $P(\mathcal{B}) = 0$ for all $\mathcal{B}$ with $\mathcal{B} \rightarrow \mathcal{A}$, then it follows $P(\mathcal{A}) = 1$. This corresponds to property 2.) of a complete labeling, see Section 2. The following proposition establishes the probabilistic analogue of the third property of a complete labeling.

**Proposition 2** *Let $P$ be p-justifiable and $\mathcal{A} \in \mathsf{Arg}$. If $P(\mathcal{A}) \in (0, 1)$ then*

1. *there is no $\mathcal{B} \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$ and $P(\mathcal{B}) = 1$ and*
2. *there is a $\mathcal{B}' \in \mathsf{Arg}$ with $\mathcal{B}' \rightarrow \mathcal{A}$ and $P(\mathcal{B}') > 0$.*

Before we investigate the relationships between our probabilistic semantics and classical argumentation semantics in more depth we analyze the properties of probabilistic semantics by itself.

**Example 5** *We continue Ex. 4. There, the probability function $P$ is p-justifiable wrt. $\mathsf{AF}$ as e. g. $P(\mathcal{A}_1) \leq 1 - P(\mathcal{A}_2)$ and $P(\mathcal{A}_4) \geq 1 - P(\mathcal{A}_3) - P(\mathcal{A}_5)$.*

The set of $p$-justifiable probability functions contains all probability functions that agree with our intuition of argumentation. This set has some nice properties as shown below.

**Proposition 3** *The set $\mathcal{P}^J_{\mathsf{AF}}$ is non-empty and convex.*

The above proposition states that for every argumentation framework $\mathsf{AF}$ there is a $p$-justifiable probability function $P$ wrt. $\mathsf{AF}$. Furthermore, the set of $p$-justifiable probability functions is closed wrt. to convex combination. That is, given two $p$-justifiable probability function $P_1$, $P_2$ and some $\delta \in [0, 1]$ it follows that $P_3$ defined via $P_3(e) = \delta P_1(e) + (1 - \delta) P_2(e)$ for each $e \in \mathcal{E}$ is also $p$-justifiable.

In order to reason with a set of probability functions one can use *model-based inductive reasoning* techniques [15], i. e., instead of reasoning with the complete set one selects some appropriate representative and performs reasoning solely based on this representative. A very important approach for that is reasoning based on the *principle of maximum entropy* [15]. For a probability function $P \in \mathcal{P}_{\mathsf{AF}}$ the *entropy* $H(P)$ of $P$ is defined as $H(P) = -\sum_{e \in \mathcal{E}} P(e) \log P(e)$ with $0 \log 0 = 0$. The entropy measures the amount of indeterminateness of a probability function $P$. A probability function $P_1$ that describes absolute certain knowledge, i. e. $P_1(e) = 1$ for some $e \in \mathcal{E}$ and $P_1(e') = 0$ for every other $e' \in \mathcal{E}$, yields minimal entropy $H(P_1) = 0$. The uniform probability function $P_0$ with $P_0(e) = 1/|\mathcal{E}|$ for every $e \in \mathcal{E}$ yields maximal entropy $H(P_0) = -\log 1/|\mathcal{E}|$.

**Definition 5** *Let $\mathcal{P} \subseteq \mathcal{P}_{\mathsf{AF}}$ be a set of probability functions.*

- *$P^* \in \mathcal{P}$ is a* maximum entropy model *of $\mathcal{P}$ if $H(P^*)$ is maximal in $\{H(P) \mid P \in \mathcal{P}\}$. Let $\mathsf{MaxE}(\mathcal{P})$ be the set of all maximum entropy models of $\mathcal{P}$.*
- *$P^* \in \mathcal{P}$ is a* minimum entropy model *of $\mathcal{P}$ if $H(P^*)$ is minimal in $\{H(P) \mid P \in \mathcal{P}\}$. Let $\mathsf{MinE}(\mathcal{P})$ be the set of all minimum entropy models of $\mathcal{P}$.*
- *$P_c$ is the* centroid *of $\mathcal{P}$ if*

$$P_c(e) = \frac{\int_{\mathcal{P}} P(e) dP(e)}{\int_{\mathcal{P}} dP(e)} \quad \text{for all } e \in \mathcal{E}.$$

A maximum entropy model $P \in \mathsf{MaxE}(\mathcal{P})$ is as unbiased as possible among the probability functions in $\mathcal{P}$, i. e., it contains as less information as possible. Reasoning based on the principle of maximum entropy is a popular approach in probabilistic reasoning as it satisfies several nice properties [15]. Here, we also consider minimum entropy models as they correspond to stable labelings (see below) and the centroid as further approaches for selecting specific models from a set of probability functions.

**Proposition 4** *If $\mathcal{P}$ is a non-empty convex set of probability functions then $|\mathsf{MaxE}(\mathcal{P})| = 1$, i. e. a maximum entropy model exists and is uniquely determined.*

For the proof of the above proposition see e. g. [15]. Taking together Propositions 3 and 4 we obtain the following nice observation as a simple corollary.

**Corollary 1** *The maximum entropy model $P^*$ of $\mathcal{P}^J_{\mathsf{AF}}$ exists and is uniquely determined.*

Note that the centroid $P_c$ of $\mathcal{P}^J_{\mathsf{AF}}$ is, by definition, also uniquely determined[2] but this is, in general, not true for minimum entropy models.

**Example 6** *We continue Ex. 4. While both the maximum entropy model $P^*$ and the centroid $P_c$ of $\mathcal{P}^J_{\mathsf{AF}}$ are uniquely determined there are three minimum entropy models $P_1^{min}, P_2^{min}, P_3^{min}$ of $\mathcal{P}^J_{\mathsf{AF}}$. The degrees of beliefs for the arguments of* AF *wrt. those models are given in Table 1, rounded to two decimal places. The maximum entropy model is as unbiased as possible, assigning a degree of belief of 0.5 to each argument, whereas the minimum entropy models have maximum information and take extreme values. The centroid $P_c$ reflects the overall situation in* AF*. For example, argument $\mathcal{A}_3$ is attacked by two arguments and receives a small degree of belief. Furthermore, both $\mathcal{A}_2$ and $\mathcal{A}_5$ each attack two other arguments and also defend themselves against attacks, therefore getting a relatively high degree of belief of 0.57 and 0.64, respectively.*

| | $P^*$ | $P_c$ | $P_1^{\min}$ | $P_2^{\min}$ | $P_3^{\min}$ |
|---|---|---|---|---|---|
| $\mathcal{A}_1$ | 0.5 | 0.43 | 0 | 1 | 0 |
| $\mathcal{A}_2$ | 0.5 | 0.57 | 1 | 0 | 1 |
| $\mathcal{A}_3$ | 0.5 | 0.14 | 0 | 0 | 0 |
| $\mathcal{A}_4$ | 0.5 | 0.36 | 1 | 0 | 0 |
| $\mathcal{A}_5$ | 0.5 | 0.64 | 0 | 1 | 1 |

**Table 1.** Degrees of belief in Ex. 6

We take a closer look on the centroid of $\mathcal{P}^J_{\mathsf{AF}}$ in the next section.

## 4 Comparison with Classical Semantics

In this section, we investigate the relationships between classical semantics and probabilistic semantics in more depth.

A probability function $P \in \mathcal{P}_{\mathsf{AF}}$ is a generalization of a labeling. For an argument $\mathcal{A} \in \mathsf{Arg}$, the probability $P(\mathcal{A}) = 1$ is equivalent to stating that the argument $\mathcal{A}$ is in and the probability $P(\mathcal{A}) = 0$ is equivalent to stating the $\mathcal{A}$ is out. A probability $P(\mathcal{A}) \in (0, 1)$ generalizes the status undec while $P(\mathcal{A}) = 0.5$ is the most "unbiased undec". Labelings can be linked to probability functions as follows. For a labeling $L$ the *characteristic probability function $P_L$* of $L$ is defined via

1. if $\mathsf{undec}(L) = \emptyset$:

$$P_L(\mathsf{in}(L)) = 1$$
$$P_L(e') = 0 \qquad \text{for all } e' \in \mathcal{E} \setminus \{\mathsf{in}(L)\}$$

2. if $\mathsf{undec}(L) \neq \emptyset$:

$$P_L(\mathsf{in}(L)) = P_L(\mathsf{in}(L) \cup \mathsf{undec}(L)) = 0.5$$
$$P_L(e') = 0 \qquad \text{for all } e' \in \mathcal{E} \setminus \{\mathsf{in}(L), \mathsf{in}(L) \cup \mathsf{undec}(L)\}$$

Note that $P_L$ is well-defined due to Prop. 1. It is easy to see, that $P_L(\mathcal{A}) = 1$ if and only if $L(\mathcal{A}) = \mathsf{in}$, $P_L(\mathcal{A}) = 0$ if and only if $L(\mathcal{A}) = \mathsf{out}$, and $P_L(\mathcal{A}) = 0.5$ if and only if $L(\mathcal{A}) = \mathsf{undec}$. Probability functions $P_1, P_2$ are *argument-equivalent*, denoted by $P_1 \equiv P_2$, if and only if $P_1(\mathcal{A}) = P_2(\mathcal{A})$ for all $\mathcal{A} \in \mathsf{Arg}$.

**Theorem 1** *Let* AF *be some abstract argumentation framework and let $L$ be some labeling.*

1. *If $L$ is complete then $P_L$ is p-justifiable.*
2. *$L$ is grounded if and only if $P_L \equiv P^*$ for $\{P^*\} = \mathsf{MaxE}(\mathcal{P}^J_{\mathsf{AF}})$.*
3. *If stable labelings exist for* AF *then $L$ is stable if and only if $P_L \in \mathsf{MinE}(\mathcal{P}^J_{\mathsf{AF}})$.*

The above theorem establishes quite interesting relationships between our probabilistic semantics and classical semantics. First, the concept or $p$-justifiable generalizes complete semantics as every complete labeling induces a $p$-justifiable probability function. Second, the grounded labeling of an argumentation framework corresponds to the maximum entropy model of all $p$-justifiable probability functions (up to argument-equivalence). Third, the set of stable labeling corresponds to the set of minimum entropy models, provided that the former set is non-empty. The final two observations link the information-theoretic concept of entropy to classical argumentation semantics. The maximum entropy model of $\mathcal{P}^J_{\mathsf{AF}}$ is the probability function which is as unbiased as possible whereas the grounded labeling is the labeling which is as cautious as possible. Furthermore, a minimum entropy model of $\mathcal{P}^J_{\mathsf{AF}}$ is a probability function that maximizes information. Similarly, a stable labeling $L$ has maximum information as it assigns to each argument either in or out.

Note that the converse of 1.) in Th. 1 does not hold in general.

[2] As $\mathcal{P}^J_{\mathsf{AF}}$ is convex it also holds that $P_c \in \mathcal{P}^J_{\mathsf{AF}}$.

**Example 7** *Consider the abstract argumentation framework* $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ *with* $\mathsf{Arg} = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ *and* $\rightarrow$ *as depicted in Fig. 2. Let* $P$ *be a probability function defined as* $P(\{\mathcal{A}_1\}) = P(\{\mathcal{A}_2\}) = 0.5$ *and* $P(e) = 0$ *for all remaining* $e \in \mathcal{E}$*. Hence* $P(\mathcal{A}_1) = P(\mathcal{A}_2) = 0.5$ *and* $P(\mathcal{A}_3) = 0$*. Note that* $P$ *is p-justifiable wrt.* $\mathsf{AF}$*. However, there is no complete labeling* $L$ *with* $L(\mathcal{A}_1) = L(\mathcal{A}_2) =$ $\mathsf{undec}$ *and* $L(\mathcal{A}_3) = \mathsf{out}$.

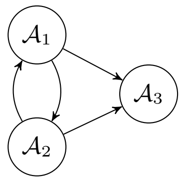


**Figure 2.** Argumentation framework from Ex. 7

Due to Th. 1 and Ex. 7 we have established that probabilistic semantics is a clear generalization of classical complete semantics. Therefore, we can integrate probabilistic semantics into the hierarchy of classical semantics as depicted in Fig. 3 (an arrow reads "is less general than").

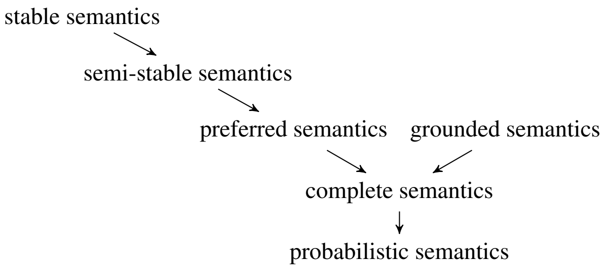


**Figure 3.** Relationships between semantics

The converse of 3.) in Th. 1 does not hold in general as well as a minimum entropy model even exists if $\mathsf{AF}$ has no stable labeling. However, the set $\mathsf{MinE}(\mathcal{P}^J_{\mathsf{AF}})$ can be characterized as follows.

**Proposition 5** *Let* $\mathsf{AF}$ *be some abstract argumentation framework and let* $L$ *be some labeling. Then* $P_L \in \mathsf{MinE}(\mathcal{P}^J_{\mathsf{AF}})$ *if and only if* $\mathsf{undec}(L)$ *is minimal wrt. set cardinality.*

Note that the above proposition does not establish that $\mathsf{MinE}(\mathcal{P}^J_{\mathsf{AF}})$ is equivalent to the set of semi-stable labelings as a semi-stable labeling is characterized by having a minimal $\mathsf{undec}(L)$ wrt. set *inclusion*. However, it holds that $L$ is a semi-stable labeling if $P_L \in \mathsf{MinE}(\mathcal{P}^J_{\mathsf{AF}})$.

In the previous section, the centroid $P_c$ of $\mathcal{P}^J_{\mathsf{AF}}$ has proven to be a good candidate for representing the set $\mathcal{P}^J_{\mathsf{AF}}$ as a whole. We now turn to its relationship with classical semantics.

**Theorem 2** *Let* $\{L_1, \ldots, L_m\}$ *be the set of complete labelings wrt.* $\mathsf{AF}$*. Then the set* $\mathcal{P}^J_{\mathsf{AF}}$ *is a polytope where* $\{P_{L_1}, \ldots, P_{L_m}\}$ *is the set of its extremal points.*

The above theorem states that the set $\mathcal{P}^J_{\mathsf{AF}}$ is the convex hull of the characteristic probability functions of all complete labelings. It also leads to a very simple characterization of the centroid $P_c$ of $\mathcal{P}^J_{\mathsf{AF}}$.

**Corollary 2** *Let* $\{L_1, \ldots, L_m\}$ *be the set of complete labelings wrt.* $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ *and let* $P_c$ *be the centroid of* $\mathcal{P}^J_{\mathsf{AF}}$*. Then*

$$P_c(\mathcal{A}) = \frac{\sum_{i=1}^{m} \delta(L_i, \mathcal{A})}{m} \quad \textit{for all } \mathcal{A} \in \mathsf{Arg}$$

*with* $\delta(L, \mathcal{A}) = 1$ *if* $L(\mathcal{A}) = \mathsf{in}$*,* $\delta(L, \mathcal{A}) = 0$ *if* $L(\mathcal{A}) = \mathsf{out}$*, and* $\delta(L, \mathcal{A}) = 0.5$ *otherwise.*

In other words, the probability of an argument in the centroid of $\mathcal{P}^J_{\mathsf{AF}}$ is its average probability with respect to all complete labelings.

## 5 Reasoning in Critical Domains

In order to illustrate the usefulness of our non-classical semantics we elaborate on an example from the medical domain. This example is inspired by an example from [14] and does not qualify for being medically accurate. Consider the argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ with $\mathsf{Arg} = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{T}_1, \mathcal{T}_2, \mathcal{C}, \mathcal{P}_1, \mathcal{P}_2\}$ and $\rightarrow$ as depicted in Fig. 4. In $\mathsf{AF}$, the arguments $\mathcal{D}_1$ and $\mathcal{D}_2$ are arguments for treating a patient suffering from blood clotting with aspirin and chlopidogrel, respectively. Both arguments attack each other as only one drug may be selected for treatment. The arguments $\mathcal{T}_1$ and $\mathcal{T}_2$ represent contradictory medical trials stating that aspirin is more effective than chlopidogrel ($\mathcal{T}_2$) and that chlopidogrel is more effective than aspirin ($\mathcal{T}_1$). Argument $\mathcal{C}$ states that chlopidogrel is too costly and should not be prescribed. Arguments $\mathcal{P}_1$ and $\mathcal{P}_2$ represent differing views on the importance of the features "health" and "low expenses": $\mathcal{P}_1$ states that the health of a patient if more important than expenses, therefore attacking argument $\mathcal{C}$. The argument $\mathcal{P}_2$ states that having low expenses is more important than a patient's health. It can easily be seen that the grounded labeling of $\mathsf{AF}$ declares each argument as $\mathsf{undec}$. Therefore, for both credulous and skeptical inference no arguments can be established as ultimately accepted. For complete, preferred, semi-stable, stable semantics several labelings can be identified, some of the declaring $\mathcal{D}_1$ as $\mathsf{in}$ and some declaring $\mathcal{D}_2$ as $\mathsf{in}$. However, for all those classical semantics each argument in $\mathsf{AF}$ can be credulously inferred and none can be skeptically inferred. The centroid $P_c$ assigns to each argument except $\mathcal{D}_2$ a degree of belief of $0.5$. The degree of belief of $\mathcal{D}_2$ is approximately $0.278$. This assignment reflects the overall situation in $\mathsf{AF}$ as $\mathcal{D}_2$ is more controversial than $\mathcal{D}_1$ due to the former's cost. Although the degree of belief in $\mathcal{D}_1$ is not very high it is still higher than $\mathcal{D}_2$ which makes $\mathcal{D}_1$ a better recommendation. Furthermore, the centroid $P_c$ also gives a concise overview on the uncertainty inherent in $\mathsf{AF}$ which supports the user in assessing his confidence when selecting a specific action.

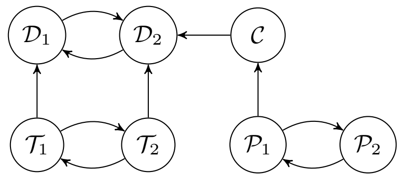


**Figure 4.** Argumentation framework for the medical domain

## 6 Related Work

The original definition of argumentation semantics by Dung [5] relies on the concept of an extension with a clear understanding of the status of an argument: an argument is either *in* an extension or not. Argument labelings [19] generalize this view and make the (already implicitly existent) third status of an argument explicit by distinguishing between arguments that are *out* and arguments that are *undecided*. Our approach generalizes this idea even further by considering the whole interval $[0, 1]$ as the space for the status of an argument. As discussed in Section 4 the classical notions of in and out can be identified with the probabilities of 1 and 0, respectively, while the argument status undec corresponds to the whole open interval $(0, 1)$, with $0.5$ being the most "unbiased" notion of undec.

To the best of our knowledge, the work reported here is the first that defines a probabilistic semantics for pure abstract argumentation frameworks. However, there are some works that extend abstract argumentation frameworks to incorporate some form of quantitative uncertainty, see e. g. [8, 6, 12, 7]. For example, the work [12] defines a probabilistic argumentation framework PAF via $\mathsf{PAF} = (\mathsf{Arg}, P_{\mathsf{Arg}}, \rightarrow, P_{\rightarrow})$ where $(\mathsf{Arg}, \rightarrow)$ is an abstract argumentation framework, $P_{\mathsf{Arg}}$ is a probability function on Arg, and $P_{\rightarrow}$ is a probability function on $\rightarrow$. A probabilistic argumentation framework PAF serves as a template for a set of abstract argumentation frameworks $\mathsf{AF}_1, \ldots, \mathsf{AF}_n$. Each $\mathsf{AF}_i$ $(i = 1, \ldots, n)$ is a sub-framework of $(\mathsf{Arg}, \rightarrow)$ and has an associated probability $P(\mathsf{AF}_i)$ of its "occurrence" which is determined by the probabilities of arguments and attacks. By fixing a specific classical semantics, e. g. grounded semantics, in [12] a probabilistic interpretation $P(\mathcal{A})$ for an argument $\mathcal{A}$ is computed by summing up the probabilities of those $\mathsf{AF}_i$ in which $\mathcal{A}$ is in the grounded extension. Similarly, the work [8] extends abstract argumentation frameworks by allowing the attack relation $\rightarrow$ to be a *fuzzy relation*. Weighted argument systems [7] assign to each attack a positive real-value to represent its strength. Reasoning in weighted argument systems is performed by fixing some threshold $\beta$ and focusing on those subsets of a system that neglects attacks with weights that sum up to at most $\beta$. The main difference between our approach and the approaches discussed so far is that they introduce additional uncertainty into the knowledge representation formalism while we assess the inherent uncertainty within abstract argumentation frameworks by a generalized semantics. A common ground of our approach and the approaches above is the focus on abstract argumentation frameworks and, therefore, the non-observance of uncertainty *within* the structure of arguments. There are also a few works that consider quantitative uncertainty within argument construction, see e. g. [11, 10, 1]. In those works additional uncertainty is introduced by weighting formulas used for creating arguments.

Similarly to our approach, the work [13] also assigns degrees of strength to arguments of an abstract argumentation framework solely based on the framework's inherent uncertainty. In [13], an argumentation framework is interpreted within an argumentation dialogue and strengths indicate how defendable an argument is for a participant. In contrast to our work, the *degrees of acceptance* in [13] have no probabilistic interpretation and are computed in a propriety way as to reflect the situation of a competitive argumentation game. Furthermore, [13] is not concerned with semantical issues of abstract argumentation frameworks.

## 7 Summary and Discussion

In this paper we proposed a new way for giving semantics to abstract argumentation frameworks. Instead of extensions or labelings we used probability functions to assign degrees of belief to arguments. We proposed a generalization of complete semantics and showed several interesting relationships between probabilistic and classical semantics on the one hand and abstract argumentation and probabilistic reasoning on the hand. In particular, we showed that the maximum entropy model of probabilistic reasoning corresponds to the grounded labeling in abstract argumentation. We also illustrated the usefulness of our approach in critical domains.

Probabilistic semantics generalizes the classical extension- and labeling-based semantics for abstract argumentation and allows for a more fine-grained differentiation of the status of arguments. For future work, we intend to investigate the relationship of probabilistic semantics with the notion of *accrual* [17] which is concerned with effects of multiple arguments attacking another argument. Roughly, it is rational to assume that the more reasons there are against a single claim the less this claim is believed to be true. In our framework, accrual of arguments is already weakly adhered for by property 2.) of Def. 4 where, in particular, the number of attacks on an argument influences the lower bound for the degree of belief in that argument, see also Ex. 7. However, a deeper analysis of this issue is left for future work.

**Acknowledgements.** The research reported here was partially supported by the SocialSensor FP7 project (EC under contract number 287975).

## REFERENCES


[1] T. Alsinet, C. I. Chesñevar, L. Godo, and G. R. Simari, 'A logic programming framework for possibilistic argumentation: Formalization and logical properties', *Fuzzy Sets and Systems*, **159**(10), 1208–1228, (2008).

[2] P. Baroni, P. E. Dunne, and M. Giacomin, 'On extension counting problems in argumentation frameworks', in *Proc. of the Third Int. Conf. on Computational Models of Argument (COMMA'10)*, pp. 63–74, (2010).

[3] P. Baroni, M. Giacomin, and G. Guida, 'Scc-recursiveness: a general schema for argumentation semantics', *Artificial Intelligence*, **168**(1–2), 162–210, (2005).

[4] M. Caminada, 'Semi-stable semantics', in *Proc. of the First Int. Conf. on Computational Models of Argument (COMMA'06)*, pp. 121–130, (2006).

[5] P. M. Dung, 'On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games', *Artificial Intelligence*, **77**(2), 321–358, (1995).

[6] P. M. Dung and P. M. Thang, 'Towards (probabilistic) argumentation for jury-based dispute resolution', in *Proc. of the Third Int. Conf. on Computational Models of Argument (COMMA'10)*, pp. 171–182, (2010).

[7] P. E. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge, 'Weighted argument systems: Basic definitions, algorithms, and complexity results', *Artificial Intelligence*, **175**(2), 457–486, (2011).

[8] J. Janssen, M. D. Cock, and D. Vermeir, 'Fuzzy argumentation frameworks', in *Proc. of the 12th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'08)*, pp. 513–520, (2008).

[9] Edwin T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[10] J. Kohlas, 'Probabilistic argumentation systems: A new way to combine logic with probability', *Journal of Applied Logic*, **1**(3-4), 225–253, (2003).

[11] P. Krause, S. Ambler, M. Elvang-Goransson, and J. Fox, 'A logic of argumentation for reasoning under uncertainty', *Computational Intelligence*, **11**(1), 113–131, (1995).

[12] H. Li, N. Oren, and T. J. Norman, 'Probabilistic argumentation frameworks', in *Proc. of the First International Workshop on the Theory and Applications of Formal Argumentation (TAFA'11)*, (2011).

[13] P.-A. Matt and F. Toni, 'Game-Theoretic Measure of Argument Strength for Abstract Argumentation', in *Proc. of the 11th European Conf. on Logics in Artificial Intelligence (JELIA'08)*, (2008).

[14] Sanjay Modgil, 'Labellings and Games for Extended Argumentation Frameworks', in *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09)*, pp. 873–878, (2009).

[15] Jeff B. Paris, *The Uncertain Reasoner's Companion – A Mathematical Perspective*, Cambridge University Press, 2006.

[16] Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1998.

[17] H. Prakken, 'A Study of Accrual of Arguments, with Applications to Evidential Reasoning', in *Proc. of the Tenth Int. Conf. on Artificial Intelligence and Law*, pp. 85–94, (2005).

[18] *Argumentation in Artificial Intelligence*, eds., I. Rahwan and G. R. Simari, Springer-Verlag, 2009.

[19] Y. Wu and M. Caminada, 'A labelling-based justification status of arguments', *Studies in Logic*, **3**(4), 12–29, (2010).