Acta Cybernetica **00** (0000) 1–17.

Sound Over-Approximation of Probabilities*

Eugenio Moggi[a], Walid Taha[b], and Johan Thunberg[b]

Abstract

Safety analysis of high confidence systems requires guaranteed bounds on the probabilities of events of interest. Establishing the correctness of algorithms that aim to compute such bounds is challenging. We address this problem in three steps. First, we use monadic transition systems (MTS) in the category of sets as a framework for modeling discrete time systems. MTS can capture different types of system behaviors, but we focus on a combination of non-deterministic and probabilistic behaviors that often arises when modeling complex systems. Second, we use the category of posets and monotonic maps as a setting to define and compare approximations. In particular, for the MTS of interest, we consider approximations of their configurations based on complete lattices. Third, by restricting to finite lattices, we obtain algorithms that compute over-approximations, i.e., bounds from above within some partial order of approximants, of the system configuration after n steps. Interestingly, finite lattices of "interval probabilities" may fail to accurately approximate configurations that are both non-deterministic and probabilistic, even for deterministic (and continuous) system dynamics. However, better choices of finite lattices are available.

Keywords: probabilities, approximation, intervals, monads

1 Introduction

Model-based safety analysis of high-confidence systems requires guaranteed bounds on the probabilities of events such as success or failure in performing a given task. Guaranteed probability bounds are needed when we consider safety or reliability of almost any real-world systems, whether it is a physical, computational, communication system, or a combination of such systems, like Cyber-Physical or Internet of Things systems.

The motivation for this work comes from an industrial collaboration where it is of interest to quantify the probability of collisions between road vehicles. In particular two vehicles approaching a crossing from two different roads. We use this

*This research was supported by the Swedish Knowledge Foundation, the ELLIIT strategic research environment, and NSF CPS project #1136099.

[a]DIBRIS, Genova University, Genova, Italy, E-mail: moggi@unige.it

[b]Halmstad University, Halmstad, Sweden, E-mail: {walid.taha, johan.thunberg}@hh.se

DOI: 10.14232/actacyb.00.0.0000.?

scenario as a motivation for the general framework we introduce. A first approach would be to model the vehicles as point masses whose time evolutions are described by a second order deterministic linear dynamic model. With such a model, one could use the explicit solutions of the dynamical systems to e.g., compute feasible regions for communication time-delays. On top of this, non-determinism as well as probabilities can be added step by step. However, such modeling and analysis does not capture the complexity of the system we like to model; it does not treat non-determinism, probabilities and physics concurrently; and in most cases it does not succeed to provide safe guarantees about the collision probabilities for the cars. We need to address this by proposing a more realistic model on the one hand and a rigorous computational framework on the other.

A more realistic model assumes two vehicles under Ackermann steering [10] and includes shapes and nonlinear kinematics and dynamics, since Ackermann steering is modeled by a dynamic system involving trigonometric expressions, making it nonlinear. Such systems do not, in general, have solutions of closed form, which means that the simplified approach described above does not apply. Furthermore, complexity is added to the model by assuming that it provides a trajectory and digital controller for the first vehicle; cooperative driving messages from the first to the second vehicle over a fading channel; a digital controller for the second vehicle to follow the first one based on the received messages, boundaries for the roads; and initial conditions. The event of concern is a *collision*, defined as a non-empty intersection between two vehicle shapes (or a vehicle shape and a road boundary).

An aspect of this type of scenario is under-specification, or *non-determinism*. It can arise in physical models of systems where multiple behaviors are possible (for example, a perfectly inverted pendulum, where two paths are possible from the initial state), in control components due to drift in clock speed, and in communication components due to unquantified external factors. In contrast to probability, non-determinism models what is not known, such as the behavior of the environment, or the actions of an opponent. Unfortunately, defining (and correctly computing) probability bounds becomes even more challenging when there is non-determinism. It is useful here to point out and distinguish the following technical problems:

1. Probabilities are real-valued quantities, that are rarely known exactly.

2. Non-determinism and probabilistic behavior are distinct features, and how to combine them correctly is not obvious.

The first problem has been addressed by extending interval methods to handle uncertainty and imprecision in probabilities (see [14]). The second problem has been addressed in the context of automata (see [13]), in order to define and verify the correctness of randomized algorithms, like those used in communication protocols.

In this paper we build on these works to establish correct ways of computing (over-approximations of) such probabilities.

1.1 Summary and Contributions

Here we provide a summary of the paper and enlist its contributions. In order to describe these contributions in the most conceivable way, we have to use some technical notation. Readers unfamiliar with the notation are referred to the respective sections under consideration for definitions and explanations.

Sec 2 recalls the definition of monad and related notions, and gives a systematic way to add non-determinism to any monad on the category of sets, in particular one can apply it to the monad of discrete probability distributions.

Sec 3 proposes monadic transition systems (MTS), which specializes the co-algebraic framework for discrete time system modeling (see [12]). A co-algebra $\alpha : S \to B(S)$, for an endofunctor B on the category of sets, describes the one-step *behavior* of a system with state space S. In a MTS B is replaced by a monad M. A monad has an underlying endofunctor, moreover allows to extend $\alpha : S \to M(S)$ to a map $\alpha^* : M(S) \to M(S)$, and view $M(S)$ as the set of *configurations*. We exemply the *discretizations* involved in the MTS modeling of continuous systems.

Sec 4 defines several complete lattices of approximants for subsets of probability distributions on a measurable space (in this context an *event* is a measurable subset of the space). These lattices are related to the notion of interval probability, which has been used for modeling uncertainty and imprecision in probabilities (see [14]).

Finally, Sec 5 applies the results in Sec 4 to define an algorithm computing over-approximations (bounds from above within a partial order of approximants) for the configuration reached after n steps by an MTS combining non-determinism and probability. We show that interval probabilities may not provide accurate over-approximations for these configurations (but other approximants do).

2 Monads

This section introduces monads, which we will use to provide a uniform treatment for non-determinism and probability. By exploiting the axiom of choice (AC), we show that non-determinism can be added to any given monad on **Set** by mere composition. This gives a way to build more complex monads, in particular to combine non-determinism and probability, while sub-monads allow to define simpler monads from existing ones.

Monads are an important notion in Category Theory [1, 2]. Moggi [7] proposes strong monads to model computational types, and Manes [6] proposes collection monads on **Set** to model collection types. We recall the definition of monad (aka Kleisli triple [5]) on **Set**, i.e., the category with sets as objects and maps (aka functions) as arrows.

Definition 2.1 (Monad)**.** *A monad on* **Set** *is a triple* $(M, \eta, \_^*)$ *such that if* X *is a set (notation* $X : \mathbf{Set}$*) then* $M(X) : \mathbf{Set}$ *and* η_X *is a map from* X *to* $M(X)$ *(notation* $\eta_X : X \to M(X)$*) called* ***unit****, and if* $f : X \to M(Y)$ *then* $f^* : M(X) \to M(Y)$ *is its* ***monadic extension****. Moreover,* η *and* $\_^*$ *satisfy the following equations for any* $f : X \to M(Y)$ *and* $g : Y \to M(Z)$

1. $f^* \circ \eta_X = f$, *namely* f^* *is an extension of* f
2. $\eta^*_X = id_{M(X)}$, *the extension of* η *is the identity*
3. $g^* \circ f^* = (g^* \circ f)^*$, *the composition of two extensions is an extension.*

We define the M*-extension of* $f : X \to Y$ *as* $M(f) = (\eta_Y \circ f)^* : M(X) \to M(Y)$.

Example 2.1. Trivial monads are the identity $I(X) = X$ and the terminal monad $1(X) = 1$, where 1 is a singleton set. Monads relevant for this work are:

Error (having one trap state) $E(X) = X + 1$, where $+$ is disjoint union, we write $ok(x)$ for an element in the left component of $X + 1$ and $fail$ for the unique element in the right component; $\eta_X(x) = ok(x)$, $f^*(ok(x)) = f(x)$ and $f^*(fail) = fail$.

Powerset (non-determinism with deadlock) $P(X)$, where $P(X)$ is the set of subsets of X, $\eta_X(x) = \{x\}$, and $f^*(A) = \bigcup_{x:A} f(x)$. Traditionally, the empty set $\emptyset$ represents deadlock (i.e., the lack of choice), and $f^*(\emptyset) = \emptyset$, since f^* preserves arbitrary unions.

The P-extension $P(f) : P(X) \to P(Y)$ of a map $f : X \to Y$, is usually called its natural set extension.

Probabilities $D_d(X) = \{p : X \to [0,1] | \sum_{x:X} p(x) = 1\}$ is the set of discrete probability distributions on X, note that the **support** $s_X(p) = \{x | p(x) > 0\}$ of p must be countable (i.e., with cardinality at most $\aleph_0$) when $\sum_{x:X} p(x)$ is bounded, $\eta_X(x)(x') = 1$ if $x = x'$ else 0, and $f^*(p)(y) = \sum_{x:X} p(x) * f(x)(y)$.

More examples, from programming language semantics, can be found in [7].

In general, monads do not compose. However, there are two ways to define monads from other monads: sub-monads and monad transformers. We recall the notion of monad map (and sub-monad), and show that the error monad E and a sub-monad P_+ of P yield two monad transformers, that map a monad M to the monads $M \circ E$ and $P_+ \circ M$, respectively.

Definition 2.2 (Monad map)**.** *A monad map from* M *to* M'*, notation* $\sigma : M \to M'$*, is a family of maps* $\sigma_X : M(X) \to M'(X)$ *indexed by* $X :$ **Set** *such that*

$$\eta'_X(x) = \sigma_X(\eta_X(x)) \qquad (\sigma_Y \circ f)^{*'}(\sigma_X(c)) = \sigma_Y(f^*(c))$$

We write Mon *for the category of monads (as objects) and monad maps (as arrows). We say that* M *is a* ***sub-monad*** *of* M' *when* $M(X) \subseteq M'(X)$ *for every* $X :$ **Set** *and the family of these inclusions is a monad map.*

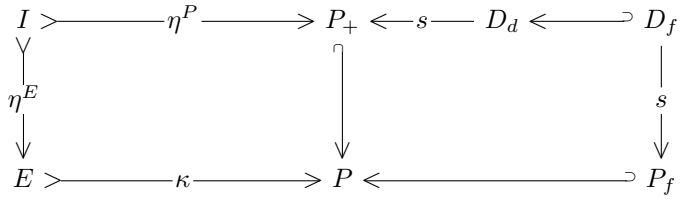
Example 2.2. The identity monad I is initial in Mon and $\eta^M : I \to M$ is the unique monad map from I to M (similarly the terminal monad 1 is terminal in Mon). In general, if M' is a monad and $\sigma_X : M(X) \to M'(X)$ is a family of injective maps, then there is at most one monad structure on M (i.e., η and $\_^*$), which makes σ a monad map. Examples of sub-monads of P and D_d are:

Non-empty powerset (non-determinism) $P_+(X)$ the set of non-empty subsets of X, equivalently $P_+(X)$ is $P(X)$ without the empty set (deadlock).

Finite powerset $P_f(X) \subseteq P(X)$ is the set of finite subsets of X.

Finite probabilities $D_f(X) \subseteq D_d(X)$ is the set of discrete probability distributions with finite support, i.e., the set of $p : D_d(X)$ such that $s_X(p)$ is finite.

Examples of monad maps relating the monads introduced so far are



where $\eta^P_X(x) = \{x\}$, $\eta^E_X(x) = ok(x)$, $\kappa_X(ok(x)) = \{x\}$ and $\kappa_X(fail) = \emptyset$. The support map $s_X : D_d(X) \to P_+(X)$ is surjective when X is a countable set.

Prop 2.1 and 2.2 show that two constructions, relevant to the rest of the paper, are monads transformers. Prop 2.1 is an instance of a well-known monad transformer, definable in any category with coproducts, while Prop 2.2 defines a new monad transformer, that is specific to the category of sets.

Proposition 2.1. *If $(M, \eta, \_^*)$ is a monad, then $M \circ E$ is the monad $(M', \eta', \_^{*'})$ defined as follows:*

- $M'(X) = M(E(X))$
- $\eta'_X(x) = \eta_X(ok(x))$
- $f^{*'}(c) = f'^*(c)$, *where* $f : X \to M(E(Y))$ *and* $f' : E(X) \to M(E(Y))$ *is the unique map such that* $f'(ok(x)) = f(x)$ *and* $f'(fail) = \eta_Y(fail)$.

Proof.

- If $f : X \to M(E(Y))$, then $f^{*'}(\eta'_X(x)) = f'(ok(x)) = f(x)$.
- ${\eta'_X}^{*'}(c) = \eta^*_{E(X)}(c) = c$.
- If $f : X \to M(E(Y))$ and $g : Y \to M(E(Z))$, then
  - $g^{*'}(f^{*'}(c)) = g'^*(f'^*(c)) = (g'^* \circ f')^*(c)$ and
  - $(g^{*'} \circ f)^{*'}(c) = (g'^* \circ f)'^*(c)$.

  Therefore, it suffices to show that $g'^* \circ f' = (g'^* \circ f)' : E(X) \to M(E(Z))$. This can be proved by case analysis on $E(X)$:
  - $g'^*(f'(ok(x)) = g'^*(f(x)) = (g'^* \circ f)(x) = (g'^* \circ f)'(ok(x))$

– $g'^*(f'(fail)) = g'^*(\eta_Y(fail)) = g'(fail) = \eta_Z(fail) = (g'^* \circ f)'(fail)$.

□

The result that $P_+ \circ M$ is a monad relies on the Axiom of Choice (**AC**):

$$\forall x : X.\exists y : Y.R(x,y) \implies \exists f : X \to Y.\forall x : X.R(x, f(x)).$$

Moreover, the result fails if P_+ is replaced by P.

Proposition 2.2. *If $(M, \eta, \_^*)$ is a monad, then $P_+ \circ M$ is the monad $(M', \eta', \_^{*'})$ defined as follows:*

- $M'(X) = P_+(M(X))$
- $\eta'_X(x) = \{\eta_X(x)\}$
- $F^{*'}(C) = \{f^*(c) | c : C \wedge f : \Pi x : X.F(x)\}$, *where* $F : X \to P_+(M(Y))$.

Proof. Given $F : X \to P_+(M(Y))$ the dependent product $\Pi x : X.F(x)$ denotes the set of maps $f : X \to M(Y)$ such that $\forall x : X.f(x) : F(x)$.

- If $F : X \to P_+(M(Y))$, then $F^{*'}(\eta'_X(x)) = \{f(x) | f : \Pi x : X.F(x)\}$.

  By **AC** exists a map $f : \Pi x : X.F(x)$, because $\forall x : X.\exists c : M(Y).c : F(x)$[1].

  Moreover, for every $x : X$ and $c : F(x)$ is in $\Pi x : X.F(x)$ also the map $f[x \mapsto c]$, which maps x to c and is equal to f on the other elements of X.

  Therefore, $\{f(x) | f : \Pi x : X.F(x)\} = F(x)$.

- ${\eta'_X}^{*'}(C) = \{\eta^*_X(c) | c : C\} = C$, since $\Pi x : X.\eta'_X(x) = \{\eta_X\}$.

- If $F : X \to P_+(M(Y))$ and $G : Y \to P_+(M(Z))$, then

  – $G^{*'}(F^{*'}(C)) = \{g^*(f^*(c)) | c : C \wedge f : \Pi x : X.F(x) \wedge g : \Pi y : Y.G(y)\}$ and

  – $(G^{*'} \circ F)^{*'}(C) = \{h^*(c) | c : C \wedge h : \Pi x : X.\{g^*(c_x) | c_x : F(x) \wedge g : \Pi y : Y.G(y)\}\}$

  $G^{*'}(F^{*'}(C)) \subseteq (G^{*'} \circ F)^{*'}(C)$ by taking $h = g^* \circ f$ and $c_x = f(x)$ for $x : X$. For the other inclusion we apply **AC** to $\forall x : X.\exists c_x : M(Y).c_x : F(x) \wedge h(x) = g^*(c_x)$ to get a map $f : X \to M(Y)$, which chooses one c_x for each $x : X$.

□

There is also a sub-monad relation between the original monad and the monad constructed by these two monad transformers.

[1] This is valid only for P_+ and not P, where it is false for every $F : X \to P(M(Y))$ such that $\exists x : X.F(x) = \emptyset \wedge \exists x : X.F(x) \neq \emptyset$.

Proposition 2.3. *The following monad maps show that (up to isomorphisms) P is a sub-monad of $P_+ \circ E$ and every monad M is a sub-monad of $P_+ \circ M$ and $M \circ E$*

$$P_+ \circ E \xleftarrow{\quad\sigma\quad} P$$

$$M \circ E \xleftarrow{\quad M(\eta^E)\quad} M \xrightarrow{\quad\eta^P\quad} P_+ \circ M$$

where $\sigma_X(A) = \{fail\}$ if $A = \emptyset$ else $\{ok(x) | x : A\}$.

3 Monadic Transition Systems

This section introduces the concept of Monadic Transition Systems (MTS), which unifies a wide range of models, including deterministic automata, non-deterministic automata, Markov chains, and probabilistic automata. At the end we exemplify the use of MTS to model a scenario related to the one described in the introduction.

A *transition system* (TS) is a pair (S, R) with R binary relation on the set S. A TS models the dynamics of a closed system, and R allows to model also the part of the closed system that we do not control, typically the environment. There is a bijection between relations $R : P(S^2)$ and maps $t : S \to P(S)$. This suggests a generalization of TS obtained by replacing the monad P with a monad M.

Definition 3.1 (Monadic TS)**.** *Given a monad $(M, \eta, \_^*)$, an M-TS is a map $t : S \to M(S)$, and we define the map $T : \mathbb{N} \to M(S) \to M(S)$ such that $T_0(c) = c$ and $T_{n+1}(c) = t^*(T_n(c))$, which gives the* configuration $T_n(c)$ *reached by the system after n steps starting from an initial configuration c.*

Example 3.1. A suitable choice of monad allows us to capture several types of computational models (where A is a set representing an input alphabet):

- Deterministic automata $t : S \to S^A$;
- Non-deterministic automata $t : S \to P(S)^A$;
- Discrete time Markov chains $t : S \to D_d(S)$;
- Probabilistic automata $t : S \to D_d(S)^A$.

The following result says that monad maps allow to view an MTS for a simpler monad as an MTS for a more complex monad.

Theorem 3.1. *If $\sigma : M \to M'$ is a monad map and $t : S \to M(S)$ is an MTS, then $t' = \sigma_S \circ t : S \to M'(S)$ is an MTS and the following diagram commutes*

$$\begin{array}{ccc} M'(S) & \xrightarrow{\quad T'_n \quad} & M'(S) \\ \uparrow{\scriptstyle \sigma_S} & & \uparrow{\scriptstyle \sigma_S} \\ M(S) & \xrightarrow{\quad T_n \quad} & M(S) \end{array}$$

thus T'_n extends T_n, when M sub-monad of M'.

Example 3.2. We explain why one should consider M-TS for M other than P.

- In P-TS one can have *deadlock states*, i.e., states s such that the set $t(s)$ of possible next states is empty. In physical systems deadlock states are not realistic, thus P_+-TS are more appropriate, as they exclude such states.
- For safety analysis it is convenient to add a $fail$ state, and add a transition from s to $fail$ when s is considered unsafe. Therefore, the appropriate choice is a $(P_+ \circ E)$-TS. Since $fail$ is a trap state, $fail : T_n(c)$ means that the system starting from the initial configuration c may fail within the first n steps.
- If a system may have also random behavior, then the appropriate choice is a $(P_+ \circ D_d \circ E)$-TS. In particular, $T_n(c)$ allows to check whether u is an upper-bound to the probability of failure within the first n steps, i.e., $\forall p : T_n(c).p(fail) \leq u$.

Our goal is the over-approximation of $T_n(c) : M(S)$. This reduces to the problem of over-approximating the monadic extension $t^* : M(S) \to M(S)$, or, more generally, a map $f : M(S) \to M(S)$. The notion of over-approximation (see Def 4.2) requires a partial order, thus we must view $M(S)$ as a subset of a partial order, and move from **Set** to the category **Po** of posets and monotonic maps (see Sec 4). When $M(S) = P_+(D_d(S+1))$ the obvious choice of complete lattice is $P(D_d(S+1))$ ordered by inclusion, where over-approximations are usually called *enclosures*.

3.1 Limitations of MTS in Set

Restricting MTS to the category **Set** of sets has benefits and limitations:

Benefits: sets are simple, every monad on **Set** is *strong* in a unique way, discrete probability distributions form a monad on **Set**, and one can add nondeterminism to any monad on **Set** (by exploiting the axiom of choice).

Limitations: sets are too simple to directly model systems with continuous time or a continuous state space S, for instance the uniform distribution on the unit interval $[0, 1]$ is not among the discrete probability distributions on $[0, 1]$.

However, there are ways to mitigate these limitations and make our results useful also for analyzing systems with continuous time, namely:

- The model of a system can be modified so that it jumps to a *trap state* (i.e., one from which the system cannot exit), when a failure occurs. This amounts to replace the state space S with $S + E$, where E is a set of trap states.
- The probability $p_t(e)$ that trap state e is reached at time t is monotone in t. Thus, we can replace continuous time with a discrete subset $\{\delta * n | n : \mathbb{N}\}$ and approximate $p_t(e)$ with an interval $[p_n(e), p_{n+1}(e)]$ when $\delta * n \leq t \leq \delta * (n+1)$.

Moreover, Sec 4 provides over-approximations for subsets of probability distributions on any *measurable space*, though in **Set** one can consider only probability distributions on *discrete spaces*.

3.2 MTS modeling of a two-car collision

As an illustration of the proposed framework, we provide an MTS-model, whose motivation stems from an industrial collaboration, where it was of interest to quantify the probability of collision between two cars approaching an intersection from two different roads. The initial configuration of the system involves non-determinism and probabilities, while the simplified deterministic dynamics models the two cars $i = 1, 2$ as point masses moving on two intersecting lines according to the ODE $x_i'(t) = v_i$, where $x_i(t)$ is the position of car i w.r.t. the intersection.

The initial positions are not known exactly, $x_i(0) : X(0) = [-15.1, -14.9]$, and the constant speeds of the two cars depend on two random variables v_i drawn from the interval $V(0) = [1.9, 2.1]$ according to the uniform distribution.

We say that a *car collision* occurs when $|x_1(t)| \leq 0.5 \wedge |x_2(t)| \leq 0.5$, i.e., when both are at most $0.5m$ from the intersection.

We take $S = ([1, 3] \times [-16, 1])^2$, where a state $((v_i, x_i)|i = 1, 2) : S$ gives speed and position of the two cars (including the speeds in the state is essential and makes the system dynamics deterministic).

We turn the above description into an MTS $f : S \to E(S)$, by replacing continuous time with discrete time (i.e., we choose a sampling interval $\delta > 0$):

fail $f((v_1, x_1), (v_2, x_2)) = fail$ when $\exists d : [0, \delta].\forall i = 1, 2.|x_i + v_i d| \leq 0.5$, else

safe $f((v_1, x_1), (v_2, x_2)) = safe$ when $\exists d : [0, \delta].\exists i = 1, 2.x_i + v_i d > 0.5$, else

move $f((v_1, x_1), (v_2, , x_2)) = ((v_1, y_1), (v_2, y_2))$ when $\forall i = 1, 2.y_i = x_i + v_i \delta \leq 1$

$fail$ is the error state added by the monad $E(-)$, and denotes a collision, while $safe$ can be any state in S such that $x_1 > 0.5 \vee x_2 > 0.5$. By composing $f : S \to E(S)$ with the monad morphism from E to $M = P_+ \circ D_d \circ E$, we can lift f to an MTS $\bar{f} : S \to M(S)$ (and use Thm 3.1), needed to handle the non-determinism in the initial positions and the random choice of accelerations.

Finally, we must define the initial configuration $c : P_+(D_d(S))$, but the uniform distribution on $V(0)$ is not discrete. Thus, we partition $V(0)$ into m intervals $V_j(0)$ of equal size, *approximate* the uniform distribution with the set of discrete distributions p such that $\forall j : m.\sum_{v:V_j(0)} p(v) = 1/m$, and define c as the set

$$\left\{ p : D_d(S) \mid \exists x_1, x_2 : X(0).\forall j, k : m. \sum_{v_1:V_j(0), v_2:V_k(0)} p((v_1, x_1), (v_2, x_2)) = \frac{1}{m^2} \right\}.$$

System vs models. We have one continuous model of the system, but a spectrum of MTS-models, with parameters δ and m: δ is for time discretization and affects only the transition map, while m affects only the initial configuration. Since these models are so simple, we can compute the exact probability of collision p_{fail} in the continuous model, and compare it with those in the MTS-models, say $p_{fail}(\delta, m)$.

The probability p_{fail} is the max for $(x_1, x_2) : X(0)^2$ of the ratio between the areas of $R_1(x_1, x_2) \cup R_2(x_1, x_2)$ and $V(0)^2$, where $R_i(x_1, x_2)$ are the convex polygons

- $R_1(x_1, x_2) = \{(v_1, v_2) : V(0)^2 | -0.5v_1 \leq x_2 v_1 + v_2(0.5 - x_1) \leq 0.5v_1\}$
- $R_2(x_1, x_2) = \{(v_1, v_2) : V(0)^2 | -0.5v_2 \leq x_1 v_2 + v_1(0.5 - x_2) \leq 0.5v_2\}$.

The max is obtained when $x_1 = x_2 = -15.9$, in this case the two polygons have the same area and disjoint interiors, thus $p_{fail} = \frac{2*|R_i(x_1,x_2)|}{0.04} = 0.86217$.

$p_{fail}(\delta, m)$ is computed similarly, but with $R_i(x_1, x_2)$ replaced by the union of the boxes in the partition of $V(0)^2$ determined by m, that intersect $R_i(x_1, x_2)$. This union does not depend on δ and $0 \leq p_{fail}(\delta, m) - p_{fail} \leq O(\frac{1}{m})$ for $m > 0$.

4 Interval Probabilities

Intervals probabilities [14] approximate probability distributions, in the same way as intervals approximate real numbers in interval arithmetic [8, 9]. In this section we address the problem of over-approximating subsets of $\Pi(X, F)$, i.e., the set of probability distributions on a measurable space (X, F). The problem is addressed by moving to the category **Po** of posets and monotonic maps, which provides the appropriate setting to define abstract interpretations [3]. We show that interval probabilities fail to accurately approximate systems behaviors that combine non-determinism and probability, even for systems as simple as that in Sec 3.2. However, there are other abstract domains, that provide more accurate approximants.

Definition 4.1 ([14])**.** *A **measurable space** is a pair (X, F), where X is a set and F is a σ-**field** (aka σ-algebra) on X, i.e., a subset $F \subseteq P(X)$ such that $\emptyset \in F$ and F is closed under complement and countable unions. $P(X)$ is the biggest σ-field on X. A **K-function** (aka probability distribution) on (X, F) is a map $\mu : F \to [0, 1]$ such that $\mu(X) = 1$ and $\mu(\cup_n A_n) = \sum_n \mu(A_n)$ for every family $(A_n | n : \mathbb{N})$ of disjoint subsets in F. We write $\Pi(X, F)$ for the set of K-functions on (X, F).*

Example 4.1. There is an injective map $\iota_X : D_d(X) \longrightarrow \Pi(X, P(X))$ given by $\iota_X(p)(A) = \sum_{x:A} p(x)$, which is bijective when X is countable. Thus, results on approximating subsets of $\Pi(X, P(X))$ turn into results on approximating subsets of $D_d(X)$. If (X, F) is a measurable space and $\mu : \Pi(X, P(X))$, then $\mu_F : \Pi(X, F)$, where $\mu_F : F \to [0, 1]$ is μ restricted to F. However, for some (X, F) there are $\mu' : \Pi(X, F)$, that are not the restriction of some $\mu : \Pi(X, P(X))$, e.g.

- If X has cardinality $\aleph_1$, then $\iota_X : D_d(X) \longrightarrow \Pi(X, P(X))$ is bijective (see [11, Thm 5.6]). Define F as the smallest σ-field on X generated by the singletons, i.e., $A : F$ if A or its complement is a countable subset of X.

  Let $\mu'(A) = 0$ if A is countable else 1, then $\mu' : \Pi(X, F)$, but μ' cannot be the restriction of some $\mu : \Pi(X, P(X))$, otherwise $\mu'(\{x\}) > 0$ for some $x : X$.

- If $X = [0, 1]$ and F is the σ-field generated by the intervals $[0, a]$ for $a : X$ (this is the σ-field generated by the standard topology on $[0, 1]$), then the *uniform distribution* $\mu' : \Pi(X, F)$ is the unique probability distribution on (X, F) such that $\mu'([0, a]) = a$.

If the *continuum hypothesis* is true, i.e., the cardinality of $[0,1]$ is $\aleph_1$, then no $\mu : \Pi(X, P(X))$ extends the uniform distribution μ' (i.e., $\mu' = \mu_F$ is false).

We use *adjunctions* to define *over-approximation* relations between two posets, a concrete domain C and an abstract domain A (in our case a poset of approximants).

Definition 4.2. *An* ***adjunction*** $\alpha \dashv \gamma$ *in* **Po** *(aka Galois connection) is a pair of maps* $C \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} A$ *in* **Po** *such that* $\forall c : |C|.\forall a : |A|.c \leq_C \gamma(a) \iff \alpha(c) \leq_A a$. *The map* γ *is called the* ***right adjoint*** *to* α, *and* α *the* ***left adjoint*** *to* γ. *We say that* $a : A$ *is an* ***over-approximation*** *of* $c : C \overset{\Delta}{\iff} c \leq_C \gamma(a)$ *(notation* $c \leq_\gamma a$*).*

Remark 4.1. A simpler definition of over-approximation, given using C only, is a is an over-approximation of c when $c \leq_C a$. However, having a separate poset A makes explicit the implementation choices about the set of *approximants*. For instance, if C is the complete lattice of subsets of $\mathbb{R}$ ordered by inclusion, possible choices for A are

1. The poset of intervals $[\underline{x}, \overline{x}]$, i.e., pairs of real numbers such that $\underline{x} \leq \overline{x}$, ordered by $[\underline{x}, \overline{x}] \leq_A [\underline{y}, \overline{y}] \overset{\Delta}{\iff} \underline{y} \leq \underline{x} \leq \overline{x} \leq \overline{y}$.
2. The finite poset of floating point intervals.
3. The finite poset of finite unions of floating point intervals.

The over-approximation relation is defined only in terms of the monotonic map γ. In the three examples of A above the definition of γ is obvious. These γ do not have a left adjoint, but it suffices to add a top $\top$ and bottom $\bot$ element and define $\gamma(\top) = \mathbb{R}$ and $\gamma(\bot) = \emptyset$, to have a left adjoint. Existence of a left adjoint α is important, since it ensures that $\alpha(c)$ is the *best* over-approximation of $c : C$, i.e., $c \leq_\gamma a \iff \alpha(c) \leq_A a$.

Definition 4.3. *The following functors allow to move between* **Set** *and* **Po**

- **Set** $\underset{\subset\!\!-J\!\!\longrightarrow}{\overset{\longleftarrow U\!\!-\!\!-}{\top}}$ **Po** $\quad \begin{array}{l} U \text{ forgetful functor } U(Y, \leq_Y) = Y \\ J \text{ embedding functor } J(X) = (X, =) \end{array} \quad J$ *left adjoint to* U
- $\mathbb{P} :$ **Set** $\longrightarrow$ **Po**, *where* $\mathbb{P}(X)$ *is the complete boolean algebra* $(P(X), \subseteq)$, $\mathbb{P}(f)$ *is the direct image map, which preserves sups (unions), thus it is monotonic.*

In **Po** our goal can be cast as follows: find adjunctions $\mathbb{P}(D) \underset{\!\!-\alpha\longrightarrow}{\overset{\longleftarrow\gamma\!\!-\!\!-}{\top}} A$, where D is a subset of $\Pi(X, F)$ with (X, F) measurable space. The goal is achieved by (Lemma 4.1 and) Thm 4.1, which offers a choice of adjunctions, where A is a complete lattice (in applications the lattices of interest are finite). Def 4.4 summarizes the poset constructions needed to define A. Thm 4.1 is proved by applying Prop 4.1, while the lemmas ensure that we are working with complete lattices

Definition 4.4 (Posets)**.** *Given two posets X and Y, one can define the posets:*

- Y^X *of monotonic maps* $\mathbf{Po}(X,Y)$ *with the point-wise order*
- $\mathbb{C}(Y)$ *of convex sets in Y, i.e.the sub-poset of* $\mathbb{P}(U(Y))$ *consisting of subsets C such that* $\forall y_1, y_2 : C. \forall y : Y. y_1 \leq_Y y \leq_Y y_2 \implies y : C$
- $\mathbb{I}(Y)$ *of intervals in Y, i.e.the sub-poset of* $\mathbb{P}(U(Y))$ *consisting of the subsets* $[\underline{y}, \overline{y}] \triangleq \{y | \underline{y} \leq_Y y \leq_Y \overline{y}\}$ *with* $\underline{y} \leq_Y \overline{y}$
- $Y_\perp$*, called lifting of Y, i.e., Y extended with a new least element* $\perp$

Any σ-field F on X is a boolean sub-algebra of the complete boolean algebra $\mathbb{P}(X)$.

Remark 4.2. It is easy to show that $\mathbb{C}(Y)$ is a complete lattice and $\mathbb{I}(Y)$ is a sub-poset of $\mathbb{C}(Y)$, moreover

- $\mathbb{C}(Y) = \mathbb{P}(Y)$, when Y is a flat poset (i.e., a set ordered by equality), and
- $\mathbb{C}(Y) \cong \mathbb{I}(Y)_\perp$, when Y is a finite linear order.

We write $[L, U]$ for the set $\{y | \exists l : L, u : U. l \leq_Y y \leq_Y u\} : \mathbb{C}(Y)$, where L and U are subsets of $U(Y)$. If Y is a finite poset, then each $C : \mathbb{C}(Y)$ is of the form $[L, U]$, where L and U are the sets of minimal and maximal elements in C, respectively.

Complete lattices, i.e., posets with all sups (and all infs), enjoy remarkable properties in relation to adjunctions. Therefore, it is useful to know under what assumptions a poset construction yields a complete lattice.

Proposition 4.1. *If X is a complete lattice and $f : \mathbf{Po}(X,Y)$, then f has a right adjoint $\iff$ f preserves sups (dually, f has a left adjoint $\iff$ f preserves infs).*

Proof. The implication from left to right is obvious, since left adjoint preserve all colimits. The other implication holds, since $f^R(y) = \sup\{x : X | f(x) \leq_Y y\}$ is a right adjoint to f, i.e., $\forall x : X. \forall y : Y. x \leq_X f^R(y) \iff f(x) \leq_Y y$, as $f(f^R(y)) \leq_Y y$. □

Lemma 4.1. *If X is a subset of Y, then* $\mathbb{P}(X) \overset{\leqslant \text{- - - } (X \cap -) \text{ - - - - } \top}{\hookrightarrow} \mathbb{P}(Y)$.

Proof. Let $\iota : \mathbf{Set}(X,Y)$ be the inclusion map, then $\mathbb{P}(\iota) : \mathbf{Po}(\mathbb{P}(X), \mathbb{P}(Y))$ is an inclusion map, which preserves sups (i.e., unions). Therefore, $\mathbb{P}(\iota)$ has a right adjoint R (by Prop 4.1), and it is immediate to check $R(B) = X \cap B$. □

Lemma 4.2. *If F is a σ-field on X (ordered by inclusion) and $[0,1]$ is the unit interval (linearly ordered), then $\Pi(X,F)$ is a subset of $U([0,1]^F)$.*

Proof. If $\mu : \mathbf{Set}(F, [0,1])$ is a probability distribution in $\Pi(X,F)$, then it is necessarily monotonic, i.e., $\mu : \mathbf{Po}(F, [0,1]) = U([0,1]^F)$. □

Lemma 4.3. *If Y is a complete lattice and X a poset, then Y^X, $\mathbb{C}(X)$ and $\mathbb{I}(Y)_\perp$ are complete lattices. Finite σ-fields on X are finite boolean sub-algebras of $\mathbb{P}(X)$.*

Proof. We prove only that $\mathbb{I}(Y)_\perp$ has infs. More precisely, we identify $\mathbb{I}(Y)_\perp$ with a subset of $\mathbb{P}(U(Y))$, namely $\perp$ is identifies with the empty set $\emptyset$, and show that it is closed under intersections computed in $\mathbb{P}(U(Y))$. Consider a subset S of $\mathbb{I}(Y)_\perp$, if $\perp : S$, then $\inf S = \perp$, otherwise $S = \{[l_i, u_i] | i : I\}$. If $\bigcap_i S = \emptyset$, then $\inf S = \perp$, otherwise $\bigcap_i S = [l, u]$ with $l \triangleq \sup_i l_i \leq u \triangleq \inf_i u_i$, i.e., $\inf S = [l, u]$. □

As stated earlier, adjunctions capture the over-approximation relation. The following theorem establishes sufficient conditions for the existence of such an adjunction for probability distributions:

Theorem 4.1 (Approximation)**.** *If (X, F) is a measurable space, F_0 is a sub-poset of F and Y_0 is a complete sub-lattice of $[0, 1]$, then there are adjunctions*

$$\mathbb{P}(\Pi(X, F)) \underset{\longrightarrow \alpha \longrightarrow}{\overset{\longleftarrow \gamma \longrightarrow}{\top}} \mathbb{I}(Y_0^{F_0})_\perp$$

where $\gamma(\perp) = \emptyset$ and $\gamma([l, u]) = \{\mu : \Pi(X, F) | \forall A : F_0. l(A) \leq \mu(A) \leq u(A)\}$.

Proof. We show that γ preserves infs for subsets $\{[l_i, u_i] | i : I\}$ of $\mathbb{I}(Y_0^{F_0})$. $Y_0^{F_0}$ is a complete lattice, thus we can define $l \triangleq \sup_i l_i$ and $u \triangleq \inf_i u_i$, then

$$\begin{aligned} \cap_i \gamma([l_i, u_i]) &= \{\mu : \Pi(X, F) | \forall i : I. \forall A : F_0. l_i(A) \leq \mu(A) \leq u_i(A)\} \\ &= \{\mu : \Pi(X, F) | \forall A : F_0. l(A) \leq \mu(A) \leq u(A)\} \\ &= \gamma([l, u]) \text{ if } (l \leq u) \text{ else } \gamma(\perp) = \gamma(\inf_i [l_i, u_i])) \end{aligned}$$

Since $\mathbb{I}(Y_0^{F_0})_\perp$ is a complete lattice, then γ has a left adjoint (by Prop 4.1). □

Remark 4.3. The adjunction in Thm 4.1 factors through $\mathbb{C}(Y_0^{F_0})$, namely

$$\mathbb{P}(\Pi(X, F)) \underset{\longrightarrow \alpha \longrightarrow}{\overset{\longleftarrow \gamma \longrightarrow}{\top}} \mathbb{C}(Y_0^{F_0}) \underset{\longrightarrow\!\!\!\!\!\!\rightarrow}{\overset{\longleftarrow\!\!\!\!\!\!\prec}{\top}} \mathbb{I}(Y_0^{F_0})_\perp.$$

Given a measurable space (X, F), we relate the notions of interval probabilities in [14] to the posets and adjunctions in Thm 4.1.

- An R-probability [14, Def 2.2] is *roughly* an interval $[l, u] : \mathbb{I}([0, 1]^F)$ such that $\gamma([l, u]) \subseteq \Pi(X, F)$ is non-empty. However, in [14] the maps $l, u : \mathbf{Set}(F, [0, 1])$ were not required by the author to be monotonic.
- An F-probability [14, Def 2.4] is an R-probability such that $[l, u] = \alpha(\gamma([l, u]))$, or equivalently $[l, u] = \alpha(D)$ for some non-empty subset D of $\Pi(X, F)$. Monotonicity of l and u is not required explicitly, but it follows from the extra axiom that a R-probability must satisfy to be a F-probability.

- A partially determined R-probability [14, Def 2.7] consists of two subsets $F_l, F_u \subseteq F$ and two maps $l : \mathbf{Set}(F_l, [0,1])$ and $u : \mathbf{Set}(F_u, [0,1])$ such that the set $\gamma([l,u]) = \{\mu : \Pi(X,F) | \forall A : F_l . l(A) \leq \mu(A) \wedge \forall A : F_u . \mu(A) \leq u(A)\}$ is non-empty. One can always extend l and u to $F_0 = F_l \cup F_u$, by taking 0 as default value for l and 1 as default value for u, so that $\gamma([l,u])$ is unchanged.

  If $F_0 \subseteq F$ is a partition of X, then $l, u : \mathbf{Set}(F_0, [0,1])$ are trivially monotonic, because the partial order on F_0 is equality.

- A partially determined F-probability [14, Def 2.8] with $F_l = F_u = F_0 \subseteq F$ is an interval $[l,u] : \mathbb{I}([0,1]^{F_0})$ such that $[l,u] = \alpha(\gamma([l,u]))$, or equivalently $[l,u] = \alpha(D)$ for some non-empty subset D of $\Pi(X,F)$.

5 Sound Over-Approximation Algorithm

Given an over-approximation relation defined by an adjunction, it is easy to specify the requirements for an algorithm computing over-approximations, and give sufficient conditions for its correctness.

Specification. Given an MTS $t : S \to M(S)$, where $M(S) = P_+(D_d(E(S)))$, and a configuration $c : M(S)$ we would like to compute $T_n(c) = (t^*)^n(c) : M(S)$. However, t and c are not suitable inputs for an algorithm, since the set $M(S)$ is uncountable (even when S is finite).

Following a standard approach in abstract interpretation, we replace $M(S)$ with a finite complete lattice A, the *abstract domain*, replace t^* and c with their *abstract interpretations* $g : \mathbf{Po}(A,A)$ and $a : A$, and compute the *abstract interpretation* $g^n(a) : A$ of $(t^*)^n(c)$.

The map t^* is monotonic and preserves non-empty unions, thus it extends in a unique way to a monotonic union preserving map $f : \mathbf{Po}(C,C)$, where C is the complete lattice $\mathbb{P}(D_d(E(S)))$, moreover $T_n(c) = f^n(c)$, since f extends t^*. In this way we have moved from **Set** to **Po**, thus we can use adjunctions in **Po** to relate the complete lattice C with an abstract domain A.

Choice of over-approximation relation. Since $D_d(E(S))$ is embedded into $\Pi(X,F)$, where $(X,F) = (E(S), P(E(S)))$, by Lemma 4.1 and Thm 4.1, any choice of finite subset F_0 of F and finite sub-lattice Y_0 of $[0,1]$ gives an adjunction between $C = \mathbb{P}(D_d(E(S)))$ and the finite lattice $A = \mathbb{I}(Y_0^{F_0})_\perp$

$$C = \mathbb{P}(D_d(E(S))) \underset{\longrightarrow \alpha \longrightarrow}{\overset{\longleftarrow \gamma \longrightarrow}{\top}} \mathbb{I}(Y_0^{F_0})_\perp = A$$

Algorithm. Given $a : A$ and $g : \mathbf{Po}(A,A)$, compute $g^n(a) : A$. Since A is a finite lattice, we have an algorithm, but we need to make some assumptions on a and g, to ensure its correctness, i.e., that $g^n(a)$ over-approximates $T_n(c)$.

Correctness. If $a : A$ over-approximates $c : C$, i.e., $c \leq_\gamma a$, and $g : \mathbf{Po}(A, A)$ over-approximates the extension f of t^*, i.e., $\forall a : A. f(\gamma(a)) \leq_\gamma g(a)$, then $g^n(a) : A$ over-approximates $T_n(c) = f^n(c) : C$, i.e., $f^n(c) \leq_\gamma g^n(a)$.

The proof is a straightforward induction on n, which relies only on having an adjunction between C and A, thus one may consider other choices of finite lattices A, besides interval probabilities. Moreover, the adjunction determines a *best choice* of over-approximations a and g, given c and f, namely: $a = \alpha(c)$ and $g = \alpha \circ f \circ \gamma$.

Accuracy. In addition to correctness one would like *accuracy*. If we focus on the probability of failure, then we can use the monotonic map $p_E : C \to [0, 1]$ mapping a configuration c to $\sup\{p(fail) | p : c\}$, and define the **inaccuracy** of the result computed by the over-approximation algorithm as $p_E(\gamma(g^n(a))) - p_E(f^n(c))$. Under the assumption of correctness, this quantity is always in the interval $[0, 1]$, and when it is closer to 0, it mean better accuracy.

The adjunction $C \underset{\longrightarrow\alpha\longrightarrow}{\overset{\longleftarrow\gamma\longleftarrow}{\top}} A$ is the critical choice for achieving accuracy, since it determines the unique *best choice* of over-approximations a and g, given the initial configuration $c : M(S)$ and the MTS $t : S \to M(S)$.

5.1 Example of Approximation

We apply the approach to the MTS of Sec 3.2, and show that the over-approximation algorithm is inaccurate, whenever the abstract domain A is of the form $\mathbb{I}(Y_0^{F_0})_\perp$. We claim (without proof) that for these simple MTS the algoritm can achieve inaccuracy less than ϵ, by choosing an abstract domain of the form $\mathbb{C}(Y_0^{F_0})$ (see Def 4.4, Rmk 4.2 and 4.3).

Recall that the state space of the MTS $f : S \to E(S)$ is $S = (S_V \times S_X)^2$, where $S_V = [1, 3]$ and $S_X = [-16, 1]$, and f depends on a sampling interval $\delta > 0$, but the probability of collision is insensitive to δ, thus we fix $\delta = 1$. The initial configuration $c : P_+(D_d(S))$ depends on a partition of $V(0) = [1.9, 2.1] \subset S_V$ into $m > 0$ intervals of size $0.2/m$, thus we write $c(m)$, to make this dependency explicit. Moreover, we must allow m to grow, in order to approximate with increasing accuracy the uniform distribution used by the continuous model. Therefore, also the abstract domain $A(m)$ should depend on m. Given $m > 0$, define $A(m) = \mathbb{I}(Y_m^{F_m})_\perp$, where

- Y_m is the sub-lattice of $[0, 1]$ whose $m + 1$ elements are i/m for $i : m + 1$,
- $F_X(m)$ is the finite partition of S_X into intervals of size $\epsilon = 0.2/m$, similarly
- $F_V(m)$ is the finite partition of S_V into intervals of size ϵ,
- F_m is the finite partition of S into hyper-cubes given by $(F_V(m) \times F_X(m))^2$.

We show that the best over-approximation $[l, u]$ of $c(m)$ in : $A(m)$ is *unsatisfactory*. To do this consider u, i.e., the smallest $u : Y_m^{F_m}$ s.t. $\forall p : c(m). \forall B : F_m. p(B) \leq u(B)$, and define a probability distribution $q \leq u$ for which the collision is certain.

1. $\forall x : X(0).\forall v : V(0).\exists p : c(m).p((v,x),(v,x)) = 1/m^2$ by definition of $c(m)$

2. for $x : X(0)$ denote with $B_X(x)$ the unique $B : F_X$ s.t. $x \in B$, similarly for $v : V(0)$ denote with $B_V(v)$ the unique $B : F_V$ s.t. $v \in B$, then

3. $\forall x : X(0).\forall v : V(0).1/m^2 \leq u((B_V(v) \times B_X(x))^2)$

4. by definition of F_X there are at least m elements in F_X that intersect $X(0)$, similarly there are at least m elements in F_V that intersect $V(0)$, denote them by $B_{X,i}$ and $B_{V,i}$ with $i : m$

5. choose two m-tuples $(x_i | i : m)$ and $(v_i | i : m)$ s.t. $\forall i : m.B_{X,i} = B_X(x_i)$ and $\forall i : m.B_{V,i} = B_V(v_i)$, and consider the probability distribution $q : D_d(S)$ s.t. $\forall i, j : m.q(s_{i,j}, s_{i,j}) = 1/m^2$, where $s_{i,j} = (v_j, x_i)$

6. clearly $\forall B : F_m.q(B) \leq u(B)$, and each s in the support of q leads to a collision, because the two cars have the same speed and position.

6 Conclusions and Future Work

The main contribution of this paper is to place the notion of interval probability in the context of the category **Po** of posets and monotonic maps (Sec 4), so that one can use general techniques from abstract interpretation to compute over-approximations of the probability of failure (Sec 5) for a system described by a monadic transition system.

Key insights from the work are the use of monads to generalize the notion of set extensions and of lattices and abstract interpretation as means for abstracting away from concrete representations of bounds. The work also raises the question of whether there is a way to avoid the reliance on the axiom of choice (AC).

Here we consider only monadic transition systems in the category of sets (Sec 2). This is fully satisfactory for modeling systems with a discrete state space, for which discrete probability distributions suffice, but not so for continuous or hybrid systems. In future work it will be interesting to explore the treatment of systems with a continuous state space. The main difficulty in replacing sets with more general *spaces* is the modeling of non-determinism. For instance, in the category of measurable spaces the Giry monad [4] plays the role of the monad D_d for modeling probabilistic systems, but there is no obvious analog of the monad P_+, and more importantly no systematic way for adding non-determinism to a monad on measurable spaces. It will be interesting to explore solutions to this problem.

Acknowledgment

We thank Eric Järpe for discussions and the reviewers for their valuable comments.

References

[1] Asperti, Andrea and Longo, Giuseppe. *Categories, Types and Scructures: an Introduction to Category Theory for the working Computer Scientist*. MIT Press, 1991.

[2] Awodey, Steve. *Category theory*. Oxford University Press, 2010.

[3] Cousot, Patrick and Cousot, Radhia. Abstract interpretation frameworks. *Journal of logic and computation*, 2(4):511–547, 1992. DOI: 10.1093/logcom/2.4.511.

[4] Giry, Michele. A categorical approach to probability theory. In *Categorical aspects of topology and analysis*, pages 68–85. Springer, 1982, 10.1007/bfb0092872.

[5] Manes, Ernest G. *Algebraic theories*. Springer, 10.1007/978-1-4612-9860-1, 1976.

[6] Manes, Ernest G. Implementing Collection Classes with Monads. *Mathematical Structures in Computer Science*, 8:231–276, 1998. DOI: 10.1017/S0960129598002515.

[7] Moggi, Eugenio. Notions of computation and monads. *Information and computation*, 93(1):55–92, 1991. DOI: 10.1016/0890-5401(91)90052-4.

[8] Moore, Ramon E. *Interval Analysis*. Prentice-Hall, 1966.

[9] Moore, Ramon E, Kearfott, R Baker, and Cloud, Michael J. *Introduction to interval analysis*. SIAM, 10.1137/1.9780898717716, 2009.

[10] Norris, William. *Modern steam road wagons*. Longmans, Green, and co., 1906.

[11] Oxtoby, John C. *Measure and category: A survey of the analogies between topological and measure spaces*, volume 2. Springer Science & Business Media, 10.1007/978-1-4684-9339-9, 2013.

[12] Rutten, Jan J.M.M. Universal coalgebra: a theory of systems. *Theoretical Computer Science*, 249(1):3–80, 2000. DOI: 10.1016/S0304-3975(00)00056-6, Modern Algebra.

[13] Segala, Roberto. *Modeling and verification of randomized distributed real-time systems*. PhD thesis, Massachusetts Institute of Technology, 1995.

[14] Weichselberger, Kurt. The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning*, 24(2):149–170, 2000. DOI: 10.1016/S0888-613X(00)00032-3.