Probabilistic Team Semantics

Jonni Virtema

Hasselt University, Belgium
jonni.virtema@gmail.com

January 17, 2019

Joint work with many people

- Arnaud Durand, Miika Hannula, Juha Kontinen, Arne Meier, and V. Approximation and Dependence via Multiteam Semantics. AMAI 2018 and FoIKS 2016.
- Arnaud Durand, Miika Hannula, Juha Kontinen, Arne Meier, and V. Probabilistic Team Semantics. FoIKS 2018.
- Miika Hannula, Åsa Hirvonen, Juha Kontinen, Vadim Kulikov, and V. Facets of Distribution Identities in Probabilistic Team Semantics. Manuscript.
- Discussions with Miika Hannula and Juha Kontinen.

Distributions

Consider:

- A collection of data from some repetitive science experiment.
- Data obtained from a poll.
- Any collection of data, that involves meaningful duplicates of data.

One natural way to represent the data is to use multisets (sets with duplicates).

Often the multiplicities themselves are not important; the distribution of data is:

- The locations of the electrons of an atom.
- Pre-election poll of party support.
- Distribution of a population with attributes like education, salary, and age.

Distributions

Consider:

- A collection of data from some repetitive science experiment.
- Data obtained from a poll.
- Any collection of data, that involves meaningful duplicates of data.

One natural way to represent the data is to use multisets (sets with duplicates).

Often the multiplicities themselves are not important; the distribution of data is:

- The locations of the electrons of an atom.
- Pre-election poll of party support.
- Distribution of a population with attributes like education, salary, and age.

Distributions

Definition

A distribution is a mapping $f : A \to [0, 1]$ from a set A of values to the closed interval $[0, 1]$ of real numbers such that the probabilities sum to 1, i.e.,

$$\sum_{a \in A} f(a) = 1.$$

- A team is a set of first-order assignments (a database without duplicates).
- A multiteam is a pair (X, m), where X is a team and $m : X \to \mathbb{N}^{>0}$ is a multiplicity function (a database with duplicates).
- A probabilistic team is a pair (X, p), where X is a team and $p : X \to [0, 1]$ is a distribution (distribution of data).

Distributions

Definition

A distribution is a mapping $f : A \to [0, 1]$ from a set A of values to the closed interval $[0, 1]$ of real numbers such that the probabilities sum to 1, i.e.,

$$\sum_{a \in A} f(a) = 1.$$

- A team is a set of first-order assignments (a database without duplicates).
- A multiteam is a pair (X, m), where X is a team and $m : X \to \mathbb{N}^{>0}$ is a multiplicity function (a database with duplicates).
- A probabilistic team is a pair (X, p), where X is a team and $p : X \to [0, 1]$ is a distribution (distribution of data).

Probabilistic teams

- Modelling of data that is inherently a probability distribution.
- Abstraction of data with duplicates.
- There is close connection between multiteams and probabilistic teams.
  - Multiteams with real number weights $\approx$ probabilistic teams.

We introduce a logic that describe properties of probabilistic teams.

We consider the expansion of first-order logic with

- marginal identity atoms $(x_1, \ldots, x_n) \approx (y_1, \ldots, y_n)$
- marginal distribution equivalence atoms $(x_1, \ldots, x_n) \approx^* (y_1, \ldots, y_n)$
- probabilistic conditional independence atoms $\overline{y} \perp\!\!\!\perp_{\overline{x}} \overline{z}$.

Probabilistic teams

- Modelling of data that is inherently a probability distribution.
- Abstraction of data with duplicates.
- There is close connection between multiteams and probabilistic teams.
  - Multiteams with real number weights $\approx$ probabilistic teams.

We introduce a logic that describe properties of probabilistic teams.

We consider the expansion of first-order logic with

- marginal identity atoms $(x_1, \ldots, x_n) \approx (y_1, \ldots, y_n)$
- marginal distribution equivalence atoms $(x_1, \ldots, x_n) \approx^* (y_1, \ldots, y_n)$
- probabilistic conditional independence atoms $\overline{y} \perp\!\!\!\perp_{\overline{x}} \overline{z}$.

Probabilistic atoms

Let $\mathbb{X} = (X, p)$ be a probablistic team and $\vec{x}, \vec{a}$ be tuples of variables and values of length k. We define

$$|\mathbb{X}|_{\vec{x}=\vec{a}} := \sum_{\substack{s \in X \\ s(\vec{x})=\vec{a}}} p(s).$$

Probabilistic atoms

Let $\mathbb{X} = (X, p)$ be a probablistic team and $\vec{x}, \vec{a}$ be tuples of variables and values of length k. We define

$$|\mathbb{X}|_{\vec{x}=\vec{a}} := \sum_{\substack{s \in X \\ s(\vec{x})=\vec{a}}} p(s).$$

Semantics for marginal identity:

$$\mathfrak{A} \models_{\mathbb{X}} \vec{x} \approx \vec{y} \quad \text{iff} \quad |\mathbb{X}|_{\vec{x}=\vec{a}} = |\mathbb{X}|_{\vec{y}=\vec{a}}, \text{ for each } \vec{a} \in A^k.$$

Probabilistic atoms

Let $\mathbb{X} = (X, p)$ be a probablistic team and $\vec{x}, \vec{a}$ be tuples of variables and values of length k. We define

$$|\mathbb{X}|_{\vec{x}=\vec{a}} := \sum_{\substack{s \in X \\ s(\vec{x})=\vec{a}}} p(s).$$

Semantics for marginal identity:

$$\mathfrak{A} \models_{\mathbb{X}} \vec{x} \approx \vec{y} \quad \text{iff} \quad |\mathbb{X}|_{\vec{x}=\vec{a}} = |\mathbb{X}|_{\vec{y}=\vec{a}}, \text{ for each } \vec{a} \in A^k.$$

Semantics for distribution equivalence:

$$\mathfrak{A} \models_{\mathbb{X}} \vec{x} \approx^* \vec{y} \quad \text{iff} \quad \{\{|\mathbb{X}|_{\vec{x}=\vec{a}} \mid \vec{a} \in A^k\}\} = \{\{|\mathbb{X}|_{\vec{y}=\vec{a}} \mid \vec{a} \in A^k\}\}.$$

Probabilistic atoms

Let $\mathbb{X} = (X, p)$ be a probablistic team and $\vec{x}, \vec{a}$ be tuples of variables and values of length k. We define

$$|\mathbb{X}|_{\vec{x}=\vec{a}} := \sum_{\substack{s \in X \\ s(\vec{x})=\vec{a}}} p(s).$$

Semantics for probabilistic marginal independence:

$\mathfrak{A} \models_{\mathbb{X}} \overline{y} \perp\!\!\!\perp_{\overline{x}} \overline{z}$ iff, for all assignments s for $\vec{x}, \vec{y}, \vec{z}$

$$|\mathbb{X}|_{\vec{x}\vec{y}=s(\vec{x}\vec{y})} \times |\mathbb{X}|_{\vec{x}\vec{z}=s(\vec{x}\vec{z})} = |\mathbb{X}|_{\vec{x}\vec{y}\vec{z}=s(\vec{x}\vec{y}\vec{z})} \times |\mathbb{X}|_{\vec{x}=s(\vec{x})}.$$

Expressing dependencies with dependencies

- Dependence atom $=\!(\vec{x}, y)$ is equivalent to the probabilistic independence atom $y \perp\!\!\!\perp_{\vec{x}} y$ and to the distribution equivalence atom $\vec{x}y \approx^* \vec{x}$.
- The atom $\vec{x} \approx^* \vec{y}$ is equivalent to the formula

$$\exists \vec{z}(=\!(\vec{y}, \vec{z}) \wedge =\!(\vec{z}, \vec{y}) \wedge \vec{x} \approx \vec{z}).$$

- Interestingly, $\vec{x} \approx \vec{y}$ is equivalent to the formula

$$\forall \overline{z}((\overline{z} \neq \overline{x} \wedge \overline{z} \neq \overline{y}) \vee ((\overline{z} = \overline{x} \vee \overline{z} = \overline{y}) \wedge \overline{z} \approx^* \overline{x} \wedge \overline{z} \approx^* \overline{y})).$$

- Finally, $\vec{x} \approx \vec{y}$ can be expressed with an $\mathrm{FO}(\perp\!\!\!\perp)$-formula.

Expressing dependencies with dependencies

- Dependence atom $=\!(\vec{x}, y)$ is equivalent to the probabilistic independence atom $y \perp\!\!\!\perp_{\vec{x}} y$ and to the distribution equivalence atom $\vec{x}y \approx^* \vec{x}$.
- The atom $\vec{x} \approx^* \vec{y}$ is equivalent to the formula

$$\exists \vec{z}\big(=\!(\vec{y}, \vec{z}) \wedge =\!(\vec{z}, \vec{y}) \wedge \vec{x} \approx \vec{z}\big).$$

- Interestingly, $\vec{x} \approx \vec{y}$ is equivalent to the formula

$$\forall \overline{z}\big((\overline{z} \neq \overline{x} \wedge \overline{z} \neq \overline{y}) \vee ((\overline{z} = \overline{x} \vee \overline{z} = \overline{y}) \wedge \overline{z} \approx^* \overline{x} \wedge \overline{z} \approx^* \overline{y})\big).$$

- Finally, $\vec{x} \approx \vec{y}$ can be expressed with an $\mathrm{FO}(\perp\!\!\!\perp)$-formula.

Expressing dependencies with dependencies

- Dependence atom $=(\vec{x}, y)$ is equivalent to the probabilistic independence atom $y \perp\!\!\!\perp_{\vec{x}} y$ and to the distribution equivalence atom $\vec{x}y \approx^* \vec{x}$.
- The atom $\vec{x} \approx^* \vec{y}$ is equivalent to the formula

$$\exists \vec{z}\big(=(\vec{y}, \vec{z}) \land =(\vec{z}, \vec{y}) \land \vec{x} \approx \vec{z}\big).$$

- Interestingly, $\vec{x} \approx \vec{y}$ is equivalent to the formula

$$\forall \overline{z}\big((\overline{z} \neq \overline{x} \land \overline{z} \neq \overline{y}) \lor ((\overline{z} = \overline{x} \lor \overline{z} = \overline{y}) \land \overline{z} \approx^* \overline{x} \land \overline{z} \approx^* \overline{y})\big).$$

- Finally, $\vec{x} \approx \vec{y}$ can be expressed with an $\mathrm{FO}(\perp\!\!\!\perp)$-formula.

Semantics of complex formulae

Definition

Let $\mathfrak{A}$ be a structure over a finite domain A, and $\mathbb{X}\colon X \to [0,1]$ a probabilistic team of $\mathfrak{A}$. The satisfaction relation $\models_{\mathbb{X}}$ for first-order logic is defined as follows:

$$\mathfrak{A} \models_{\mathbb{X}} x = y \Leftrightarrow \text{for all } s \in X : \text{if } \mathbb{X}(s) > 0, \text{ then } s(x) = s(y)$$

$$\mathfrak{A} \models_{\mathbb{X}} x \neq y \Leftrightarrow \text{for all } s \in X : \text{if } \mathbb{X}(s) > 0, \text{ then } s(x) \neq s(y)$$

$$\mathfrak{A} \models_{\mathbb{X}} R(\overline{x}) \Leftrightarrow \text{for all } s \in X : \text{if } \mathbb{X}(s) > 0, \text{ then } s(\overline{x}) \in R^{\mathfrak{A}}$$

$$\mathfrak{A} \models_{\mathbb{X}} \neg R(\overline{x}) \Leftrightarrow \text{for all } s \in X : \text{if } \mathbb{X}(s) > 0, \text{ then } s(\overline{x}) \notin R^{\mathfrak{A}}$$

$$\mathfrak{A} \models_{\mathbb{X}} (\psi \wedge \theta) \Leftrightarrow \mathfrak{A} \models_{\mathbb{X}} \psi \text{ and } \mathfrak{A} \models_{\mathbb{X}} \theta$$

Semantics of complex formulae

Definition

Let $\mathfrak{A}$ be a structure over a finite domain A, and $\mathbb{X} \colon X \to [0,1]$ a probabilistic team of $\mathfrak{A}$. The satisfaction relation $\models_{\mathbb{X}}$ for first-order logic is defined as follows:

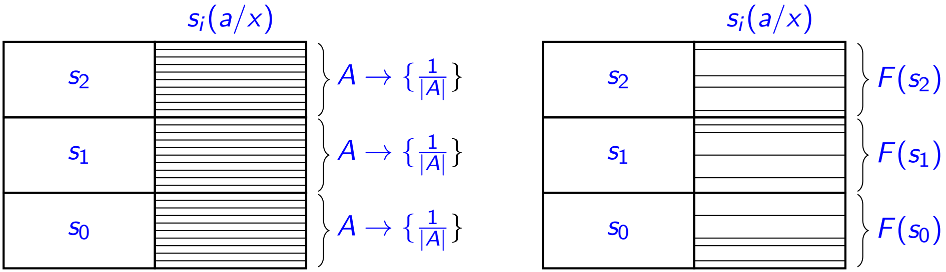
$$\mathfrak{A} \models_{\mathbb{X}} (\psi \vee \theta) \Leftrightarrow \mathfrak{A} \models_{\mathbb{Y}} \psi \text{ and } \mathfrak{A} \models_{\mathbb{Z}} \theta \text{ for some } \mathbb{Y}, \mathbb{Z} \ s.t. \ \mathbb{Y} \sqcup \mathbb{Z} = \mathbb{X}$$

$$\mathfrak{A} \models_{\mathbb{X}} \forall x \psi \Leftrightarrow \mathfrak{A} \models_{\mathbb{X}[A/x]} \psi$$

$$\mathfrak{A} \models_{\mathbb{X}} \exists x \psi \Leftrightarrow \mathfrak{A} \models_{\mathbb{X}[F/x]} \psi \text{ holds for some } F \colon X \to p_A.$$

Above p_A denote the set those distributions that have domain A.

Intuition of the quantifiers



- Universal quantification (i.e., the set $\mathbb{X}[A/x]$) is depicted on left.
- Existential quantification (i.e., the set $\mathbb{X}[F/x]$) is depicted on right.
- Height of a box corresponds to the probability of an assignment.

Intuition behind the disjunction

Question: How do we split distributions?
Answer: We rescale.

Let $\mathbb{X}\colon X \to [0,1]$ and $\mathbb{Y}\colon Y \to [0,1]$ be probabilistic teams and $k \in [0,1]$ be a real number.

We denote by $\mathbb{X} \sqcup_k \mathbb{Y}$ the k-scaled union of $\mathbb{X}$ and $\mathbb{Y}$, that is, the probabilistic team $\mathbb{X} \sqcup_k \mathbb{Y}\colon X \cup Y \to [0,1]$ defined s.t. for each $s \in X \cup Y$,

$$(\mathbb{X} \sqcup_k \mathbb{Y})(s) := \begin{cases} k \cdot \mathbb{X}(s) + (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in X \text{ and } s \in Y, \\ k \cdot \mathbb{X}(s) & \text{if } s \in X \text{ and } s \notin Y, \\ (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in Y \text{ and } s \notin X. \end{cases}$$

We then write that $Z = \mathbb{X} \sqcup \mathbb{Y}$ if $Z = \mathbb{X} \sqcup_k \mathbb{Y}$, for some k.

Intuition behind the disjunction

Question: How do we split distributions?
Answer: We rescale.

Let $\mathbb{X}\colon X \to [0,1]$ and $\mathbb{Y}\colon Y \to [0,1]$ be probabilistic teams and $k \in [0,1]$ be a real number.

We denote by $\mathbb{X} \sqcup_k \mathbb{Y}$ the k-scaled union of $\mathbb{X}$ and $\mathbb{Y}$, that is, the probabilistic team $\mathbb{X} \sqcup_k \mathbb{Y}\colon X \cup Y \to [0,1]$ defined s.t. for each $s \in X \cup Y$,

$$(\mathbb{X} \sqcup_k \mathbb{Y})(s) := \begin{cases} k \cdot \mathbb{X}(s) + (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in X \text{ and } s \in Y, \\ k \cdot \mathbb{X}(s) & \text{if } s \in X \text{ and } s \notin Y, \\ (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in Y \text{ and } s \notin X. \end{cases}$$

We then write that $Z = \mathbb{X} \sqcup \mathbb{Y}$ if $Z = \mathbb{X} \sqcup_k \mathbb{Y}$, for some k.

Intuition behind the disjunction

Question: How do we split distributions?
Answer: We rescale.

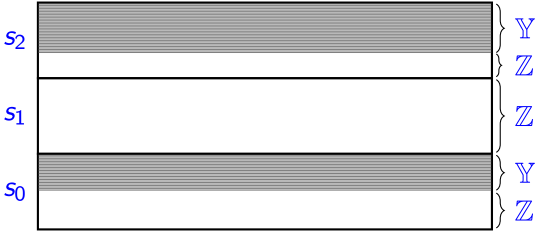
Let $\mathbb{X}\colon X \to [0,1]$ and $\mathbb{Y}\colon Y \to [0,1]$ be probabilistic teams and $k \in [0,1]$ be a real number.

We denote by $\mathbb{X} \sqcup_k \mathbb{Y}$ the k-scaled union of $\mathbb{X}$ and $\mathbb{Y}$, that is, the probabilistic team $\mathbb{X} \sqcup_k \mathbb{Y}\colon X \cup Y \to [0,1]$ defined s.t. for each $s \in X \cup Y$,

$$(\mathbb{X} \sqcup_k \mathbb{Y})(s) := \begin{cases} k \cdot \mathbb{X}(s) + (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in X \text{ and } s \in Y, \\ k \cdot \mathbb{X}(s) & \text{if } s \in X \text{ and } s \notin Y, \\ (1-k) \cdot \mathbb{Y}(s) & \text{if } s \in Y \text{ and } s \notin X. \end{cases}$$

We then write that $Z = \mathbb{X} \sqcup \mathbb{Y}$ if $Z = \mathbb{X} \sqcup_k \mathbb{Y}$, for some k.

Intuition behind the disjunction



- Partition $\mathbb{X}$ to two probabilistic teams $\mathbb{Y}$ and $\mathbb{Z}$ and re-scale both back to 1.
- **NB.** The empty set is considered as a probabilistic team.

Example

Consider a database table that lists results of experiments as a multiteam or as the related probabilistic team using the counting measure.

- Records: Outcomes of measurements obtained simultaneously in two locations.
- Attributes: Test1 and Test2 ranging over types of measurements, and Outcome1 and Outcome2 ranging over outcomes of the measurements.

The probabilistic independence atom $\mathrm{Test1} \perp\!\!\!\perp \mathrm{Test2}$ expresses that the types of measurements are independently picked in the two locations.

The marginal identity atom $(\mathrm{Test1}, \mathrm{Outcome1}) \approx (\mathrm{Test2}, \mathrm{Outcome2})$ expresses that the distributions of tests and results are the same in both test sites.

The formula $\mathrm{Test1} = \mathrm{Test2} \vee (\mathrm{Test1} \neq \mathrm{Test2} \wedge \mathrm{Outcome1} \perp\!\!\!\perp \mathrm{Outcome2})$ expresses that there is no correlation between outcomes of the different measurements.

More examples

- The formula $\forall \vec{y}\, \vec{x} \approx \vec{y}$ states that the probabilities for $\overline{x}$ are uniformly distributed over all value sequences of length $|\overline{x}|$.
- The probability of $P(x)$ is at least twice the probability of $Q(x)$.
- Can we characterise the expressive power of $\mathrm{FO}(\approx)$, $\mathrm{FO}(\perp\!\!\!\perp)$, etc., in the probabilistic setting?

More examples

- The formula $\forall \vec{y}\, \vec{x} \approx \vec{y}$ states that the probabilities for $\overline{x}$ are uniformly distributed over all value sequences of length $|\overline{x}|$.
- The probability of $P(x)$ is at least twice the probability of $Q(x)$.
- Can we characterise the expressive power of $\mathrm{FO}(\approx)$, $\mathrm{FO}(\perp\!\!\!\perp)$, etc., in the probabilistic setting?

Benchmark logic

- In team semantics context fragments of second-order logic are captured.
- $\mathrm{FO}(\perp)$ (team semantics) is as expressive as existential second-order logic.
- We define a two-sorted variant of ESO in which we allow
  - quantification of distributions, which constitute the base of numerical terms,
  - sum and multiplication on numerical terms.
- This logic characterises the expressive power of $\mathrm{FO}(\perp\!\!\!\perp)$.

Benchmark logic

- In team semantics context fragments of second-order logic are captured.
- $\mathrm{FO}(\perp)$ (team semantics) is as expressive as existential second-order logic.
- We define a two-sorted variant of ESO in which we allow
  - quantification of distributions, which constitute the base of numerical terms,
  - sum and multiplication on numerical terms.
- This logic characterises the expressive power of $\mathrm{FO}(\perp\!\!\!\perp)$.

Relative expressive power

PTS:	$\mathrm{FO}(\approx) < \mathrm{FO}(\approx, =\!(\cdot)) \equiv \mathrm{FO}(\approx^*) \leq \mathrm{FO}(\perp\!\!\!\perp) \equiv \mathrm{FO}(\perp\!\!\!\perp_c)$
TS:	$\mathrm{FO}(\subseteq) < \mathrm{FO}(\subseteq, =\!(\cdot)) \equiv \mathrm{FO}(\perp) \equiv \mathrm{FO}(\perp_c)$

Table: Expressivity in probabilistic team semantics (PTS) and team semantics (TS). Results for TS by Galliani 2012 and Galliani, Väänänen 2014.

Probabilistic structures

Definition

Let τ and σ be a relational and a functional vocabulary. A probabilistic $\tau \cup \sigma$-structure is a tuple

$$\mathfrak{A} = (A, [0,1], (R_i^{\mathfrak{A}})_{R_i \in \tau}, (f_i^{\mathfrak{A}})_{f_i \in \sigma}),$$

where

- A (i.e. the domain of $\mathfrak{A}$) is a finite nonempty set,
- $[0,1]$ is the closed interval of real numbers between 0 and 1,
- each $R_i^{\mathfrak{A}}$ is a relation on A (i.e., a subset of $A^{\mathrm{ar}(R_i)}$),
- each $f_i^{\mathfrak{A}}$ is a probability distribution from $A^{\mathrm{ar}(f_i)}$ to $[0,1]$ (i.e., a function such that $\sum_{\vec{a} \in A^{\mathrm{ar}(f_i)}} f_i(\vec{a}) = 1$).

Second-order logic for probabilistic structures

- As first-order terms we have first-order variables.
- The set of numerical σ-terms i is defined via the grammar

$$i ::= f(\vec{x}) \mid i \times i \mid \mathrm{SUM}_{\vec{x}}\, i(\vec{x}, \vec{y}),$$

where $\vec{x}, \vec{y}$ are tuples of first-order variables, $f \in \sigma$ and σ is a set of functions.

- The value of a numerical term i in a structure $\mathfrak{A}$ under an assignment s is denoted by $[i]^{\mathfrak{A}}_s$ and defined as follows:

$$[f(\overline{x})]^{\mathfrak{A}}_s := f^{\mathfrak{A}}(s(\overline{x})), \qquad [i \times j]^{\mathfrak{A}}_s := [i]^{\mathfrak{A}}_s \cdot [j]^{\mathfrak{A}}_s,$$

$$[\mathrm{SUM}_{\vec{x}}\, i(\vec{x}, \vec{y})]^{\mathfrak{A}}_s := \sum_{\vec{a} \in A^{|\vec{x}|}} [i(\vec{a}, \vec{y})]^{\mathfrak{A}}_s,$$

where $\cdot$ and $\sum$ are the multiplication and sum of real numbers.

Second-order logic for probabilistic structures

- As first-order terms we have first-order variables.
- The set of numerical σ-terms i is defined via the grammar

$$i ::= f(\vec{x}) \mid i \times i \mid \mathrm{SUM}_{\vec{x}}\, i(\vec{x}, \vec{y}),$$

where $\vec{x}, \vec{y}$ are tuples of first-order variables, $f \in \sigma$ and σ is a set of functions.

- The value of a numerical term i in a structure $\mathfrak{A}$ under an assignment s is denoted by $[i]^{\mathfrak{A}}_{s}$ and defined as follows:

$$[f(\overline{x})]^{\mathfrak{A}}_{s} := f^{\mathfrak{A}}(s(\overline{x})), \qquad [i \times j]^{\mathfrak{A}}_{s} := [i]^{\mathfrak{A}}_{s} \cdot [j]^{\mathfrak{A}}_{s},$$

$$[\mathrm{SUM}_{\vec{x}}\, i(\vec{x}, \vec{y})]^{\mathfrak{A}}_{s} := \sum_{\vec{a} \in A^{|\vec{x}|}} [i(\vec{a}, \vec{y})]^{\mathfrak{A}}_{s},$$

where $\cdot$ and $\sum$ are the multiplication and sum of real numbers.

Second-order logic for probabilistic structures

Definition

The formulae of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ is defined via the following grammar:

$$\phi ::= x = y \mid x \neq y \mid R(\vec{x}) \mid \neg R(\vec{x}) \mid i = j \mid \phi \wedge \phi \mid \phi \vee \phi \mid \exists x \phi \mid \forall x \phi \mid \exists f \phi,$$

where i is a numerical term, R is a relation symbol, f is a function variable, $\vec{x}$ is a tuple of first-order variables.

Semantics of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ is defined via probabilistic structures and assignments analogous to FO. In addition to the clauses of FO, we have:

$$\mathfrak{A} \models_s i = j \Leftrightarrow [i]^{\mathfrak{A}}_s = [j]^{\mathfrak{A}}_s,$$

$$\mathfrak{A} \models_s \exists f \phi \Leftrightarrow \mathfrak{A}[h/f] \models_s \phi \text{ for some probability distribution } h\colon A^{\mathrm{ar}(f)} \to [0,1],$$

where $\mathfrak{A}[h/f]$ denotes the expansion of $\mathfrak{A}$ that interprets f to h.

Second-order logic for probabilistic structures

Definition

The formulae of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ is defined via the following grammar:

$$\phi ::= x = y \mid x \neq y \mid R(\vec{x}) \mid \neg R(\vec{x}) \mid i = j \mid \phi \wedge \phi \mid \phi \vee \phi \mid \exists x \phi \mid \forall x \phi \mid \exists f \phi,$$

where i is a numerical term, R is a relation symbol, f is a function variable, $\vec{x}$ is a tuple of first-order variables.

Semantics of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ is defined via probabilistic structures and assignments analogous to FO. In addition to the clauses of FO, we have:

$$\mathfrak{A} \models_s i = j \Leftrightarrow [i]^{\mathfrak{A}}_s = [j]^{\mathfrak{A}}_s,$$

$$\mathfrak{A} \models_s \exists f \phi \Leftrightarrow \mathfrak{A}[h/f] \models_s \phi \text{ for some probability distribution } h \colon A^{\mathrm{ar}(f)} \to [0,1],$$

where $\mathfrak{A}[h/f]$ denotes the expansion of $\mathfrak{A}$ that interprets f to h.

Examples

- Uniformity of a distribution f can be expressed with

$$\phi(f) := \forall \overline{xy}(f(\overline{x}) = 0 \vee f(\overline{y}) = 0 \vee f(\overline{x}) = f(\overline{y})).$$

- For a numerical term i and rational number $\frac{p}{q}$, the property

$$i(\overline{x}) = \frac{p}{q}$$

can be expressed in $\mathrm{ESOf}(\mathrm{SUM}, \times)$.

Benchmark logics and probabilistic team semantics

- For a probabilistic team $\mathbb{X}\colon X \to [0,1]$, we let $f_{\mathbb{X}} : A^n \to [0,1]$ be the probability distribution that encodes $\mathbb{X}$.
- Translations are between formulae using team semantics and formulae of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ with $f_{\mathbb{X}}$ as a free variable interpreting the team.
- $\mathrm{FO}(\perp\!\!\!\perp)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM}, \times)$.
- $\mathrm{FO}(\approx^*)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM})$.
- Conjecture: $\mathrm{FO}(\approx^*) < \mathrm{FO}(\perp\!\!\!\perp)$.

Benchmark logics and probabilistic team semantics

- For a probabilistic team $\mathbb{X}\colon X \to [0,1]$, we let $f_{\mathbb{X}} : A^n \to [0,1]$ be the probability distribution that encodes $\mathbb{X}$.
- Translations are between formulae using team semantics and formulae of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ with $f_{\mathbb{X}}$ as a free variable interpreting the team.
- $\mathrm{FO}(\perp\!\!\!\perp)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM}, \times)$.
- $\mathrm{FO}(\approx^*)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM})$.
- Conjecture: $\mathrm{FO}(\approx^*) < \mathrm{FO}(\perp\!\!\!\perp)$.

Benchmark logics and probabilistic team semantics

- For a probabilistic team $\mathbb{X}\colon X \to [0,1]$, we let $f_{\mathbb{X}} : A^n \to [0,1]$ be the probability distribution that encodes $\mathbb{X}$.
- Translations are between formulae using team semantics and formulae of $\mathrm{ESOf}(\mathrm{SUM}, \times)$ with $f_{\mathbb{X}}$ as a free variable interpreting the team.
- $\mathrm{FO}(\perp\!\!\!\perp)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM}, \times)$.
- $\mathrm{FO}(\approx^*)$ is equivalent to $\mathrm{ESOf}(\mathrm{SUM})$.
- Conjecture: $\mathrm{FO}(\approx^*) < \mathrm{FO}(\perp\!\!\!\perp)$.

Relation to earlier works

Probabilistic structures are closely related to metafinite structures (Grädel, Gurevich '98), such as $\mathbb{R}$-structures (Grädel, Meer '95) that consist of a finite structure $\mathfrak{A}$ together with an ordered field of reals and a finite set of weight functions from $\mathfrak{A}$ to $\mathbb{R}$.

$\mathbb{R}$-structures can be analyzed in terms of $\mathrm{ESO}_{\mathbb{R}}(+, \times, <, (c_k)_{k\in\mathbb{R}})$, i.e., a two-sorted variant of ESO with existential quantification over functions from $\mathfrak{A}$ to reals.

Expressivity of $\mathrm{ESO}_{\mathbb{R}}(+, \times, <, (c_k)_{k\in\mathbb{R}})$ can be characterized in terms of Blum–Shub–Smale machines, i.e., a model of computation which treats real numbers as basic entities and performs arithmetic operations on reals in a single step.

Theorem (Grädel, Meer '95)

$\mathrm{ESO}_{\mathbb{R}}(+, \times, <, (c_k)_{k\in\mathbb{R}}) \equiv \mathrm{NP}_{\mathbb{R}}$, *where* $\mathrm{NP}_{\mathbb{R}}$ *is non-deterministic polynomial time over BSS machines.*

Conclusion

- Probabilistic team semantics extends team semantics by adding a probability measure over assignments.
- This makes possible to introduce logics for probabilistic dependencies such as $\perp\!\!\!\perp$ and $\approx$.
- The logics obtained can be compared to each other and characterized in terms of a two-sorted variant of ESO.
- Open problems:
  - Can we axiomatize $\mathrm{PL}(\perp\!\!\!\perp, \approx)$?
  - Data complexity of $\mathrm{FO}(\perp\!\!\!\perp), \mathrm{FO}(\approx)$? Can we logically characterize e.g. $\mathrm{P}_{\mathbb{R}}/\mathrm{NP}_{\mathbb{R}}$ classes of probability distributions in probabilistic team semantics?

Conclusion

Thanks!

- Probabilistic team semantics extends team semantics by adding a probability measure over assignments.
- This makes possible to introduce logics for probabilistic dependencies such as $\perp\!\!\!\perp$ and $\approx$.
- The logics obtained can be compared to each other and characterized in terms of a two-sorted variant of ESO.
- Open problems:
  - Can we axiomatize $\mathsf{PL}(\perp\!\!\!\perp, \approx)$?
  - Data complexity of $\mathsf{FO}(\perp\!\!\!\perp), \mathsf{FO}(\approx)$? Can we logically characterize e.g. $\mathsf{P}_{\mathbb{R}}/\mathsf{NP}_{\mathbb{R}}$ classes of probability distributions in probabilistic team semantics?