Approximation and Dependence via Multiteam Semantics

Jonni Virtema

University of Helsinki, Finland
jonni.virtema@gmail.com

Joint work with Arnaud Durand, Miika Hannula, Juha Kontinen, and Arne Meier
To appear in FoIKS 2016

10th of February, 2016

What do we do?

- Multiteam semantics: Shift from (set) teams to their multiset analogues.
- Probabilistic atoms:
  - Probabilistic inclusion atom.
  - Probabilistic conditional independence atom.
  - Probabilistic marginal independence atom.
- Basic properties of logics with the above ingredients.
- Approximate operators inspired by approximate dependence atoms by Väänänen.
- Complexity of model checking with the approximate operator.

From teams to multiteams

- *Multiset* is a pair (A, m), where A is a set and $m : A \to \mathbb{N}$ a function.
- *Team* is set X of assignments $s : VAR \to A$ with a common domain.
- Multiset (X, m) is a *multiteam* whenever X is a team.

For multisets (A, m), we define the *canonical set representative* as follows

$$[(A, m)]_{\mathrm{cset}} := \{(a, i) \mid a \in A, 0 < i \leq m(a)\}.$$

From teams to multiteams

- *Multiset* is a pair (A, m), where A is a set and $m : A \to \mathbb{N}$ a function.
- *Team* is set X of assignments $s : VAR \to A$ with a common domain.
- Multiset (X, m) is a *multiteam* whenever X is a team.

For multisets (A, m), we define the *canonical set representative* as follows

$$[(A, m)]_{\text{cset}} := \{(a, i) \mid a \in A, 0 < i \leq m(a)\}.$$

Multiteam semantics

- Replace structures by multistructures
  - Domains are multisets.
  - Relations are over the underlining set domains.
  - (The same effect as replacing identity by a equivalence relation that respects relations in the vocabulary)
- Replace teams by multiteams.
- Semantics is defined like team semantics but with canonical set representatives.

Multiteam semantics

- Replace structures by multistructures
  - Domains are multisets.
  - Relations are over the underlining set domains.
  - (The same effect as replacing identity by a equivalence relation that respects relations in the vocabulary)
- Replace teams by multiteams.
- Semantics is defined like team semantics but with canonical set representatives.

Multiteam semantics cont.

(X, m) a multiteam and (A, n) a finite multiset.

- $(A, m) \uplus (B, n)$ denotes the disjoint union of (A, m) and (B, n).
- $\mathcal{P}^+((A, m))$ is the set of non-empty submultisets of (A, m).
- For universal quntifier, define $(X, m)[(A, n)/x]$ as

$$\biguplus_{s \in X} \biguplus_{a \in A} \{(s(a/x), m(s) \cdot n(a))\}.$$

- For existential quantifier, define $X[F/x]$ as

$$\biguplus_{s \in X} \biguplus_{1 \leq i \leq m(s)} \{(s(b/x), l(b)) \mid (B, l) = F((s, i)), b \in B\},$$

where $F\colon [(X, m)]_{\text{cset}} \to \mathcal{P}^+((A, n))$ a function.

Multiteam semantics cont.

(X, m) a multiteam and (A, n) a finite multiset.

- $(A, m) \uplus (B, n)$ denotes the disjoint union of (A, m) and (B, n).
- $\mathcal{P}^+((A, m))$ is the set of non-empty submultisets of (A, m).
- For universal quntifier, define $(X, m)[(A, n)/x]$ as

$$\biguplus_{s \in X} \biguplus_{a \in A} \{(s(a/x), m(s) \cdot n(a))\}.$$

- For existential quantifier, define $X[F/x]$ as

$$\biguplus_{s \in X} \biguplus_{1 \leq i \leq m(s)} \{(s(b/x), l(b)) \mid (B, l) = F((s, i)), b \in B\},$$

where $F \colon [(X, m)]_{\mathrm{cset}} \to \mathcal{P}^+((A, n))$ a function.

Multiteam semantics cont.

(X, m) a multiteam and (A, n) a finite multiset.

- $(A, m) \uplus (B, n)$ denotes the disjoint union of (A, m) and (B, n).
- $\mathcal{P}^+((A, m))$ is the set of non-empty submultisets of (A, m).
- For universal quntifier, define $(X, m)[(A, n)/x]$ as

$$\biguplus_{s \in X} \biguplus_{a \in A} \{(s(a/x), m(s) \cdot n(a))\}.$$

- For existential quantifier, define $X[F/x]$ as

$$\biguplus_{s \in X} \biguplus_{1 \leq i \leq m(s)} \{(s(b/x), l(b)) \mid (B, l) = F((s, i)), b \in B\},$$

where $F\colon [(X, m)]_{\mathrm{cset}} \to \mathcal{P}^+((A, n))$ a function.

Multiteam semantics cont.

Definition (Multiteam semantics)

$\mathfrak{A}$ a τ-multistructure, (A, n) the domain of $\mathfrak{A}$, and (X, m) a multiteam over $\mathfrak{A}$.

$$
\begin{aligned}
\mathfrak{A} \models_{(X,m)} x = y &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) = s(y)\\
\mathfrak{A} \models_{(X,m)} x \neq y &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) \neq s(y)\\
\mathfrak{A} \models_{(X,m)} R(\vec{x}) &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \in R^{\mathfrak{A}}\\
\mathfrak{A} \models_{(X,m)} \neg R(\vec{x}) &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \notin R^{\mathfrak{A}}\\
\mathfrak{A} \models_{(X,m)} (\psi \wedge \theta) &\Leftrightarrow \mathfrak{A} \models_{(X,m)} \psi \text{ and } \mathfrak{A} \models_{(X,m)} \theta\\
\mathfrak{A} \models_{(X,m)} (\psi \vee \theta) &\Leftrightarrow \mathfrak{A} \models_{(Y,k)} \psi \text{ and } \mathfrak{A} \models_{(Z,l)} \theta \text{ for some multisets}\\
&\qquad (Y,k),(Z,l) \subseteq (X,m) \text{ s.t. } (X,m) \subseteq (Y,k) \uplus (Z,l).\\
\mathfrak{A} \models_{(X,m)} \forall x \psi &\Leftrightarrow \mathfrak{A} \models_{(X,m)[(A,n)/x]} \psi\\
\mathfrak{A} \models_{(X,m)} \exists x \psi &\Leftrightarrow \mathfrak{A} \models_{(X,m)[F/x]} \psi \text{ holds for some function}\\
&\qquad F \colon [(X,m)]_{\mathrm{cset}} \to \mathcal{P}^{+}((A,n)).
\end{aligned}
$$

Multiteam semantics cont.

Definition (Multiteam semantics)

$\mathfrak{A}$ a τ-multistructure, (A, n) the domain of $\mathfrak{A}$, and (X, m) a multiteam over $\mathfrak{A}$.

$$
\begin{aligned}
\mathfrak{A} \models_{(X,m)} x = y \quad &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) = s(y)\\
\mathfrak{A} \models_{(X,m)} x \neq y \quad &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) \neq s(y)\\
\mathfrak{A} \models_{(X,m)} R(\vec{x}) \quad &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \in R^{\mathfrak{A}}\\
\mathfrak{A} \models_{(X,m)} \neg R(\vec{x}) \quad &\Leftrightarrow \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \notin R^{\mathfrak{A}}\\
\mathfrak{A} \models_{(X,m)} (\psi \wedge \theta) &\Leftrightarrow \mathfrak{A} \models_{(X,m)} \psi \text{ and } \mathfrak{A} \models_{(X,m)} \theta\\
\mathfrak{A} \models_{(X,m)} (\psi \vee \theta) &\Leftrightarrow \mathfrak{A} \models_{(Y,k)} \psi \text{ and } \mathfrak{A} \models_{(Z,l)} \theta \text{ for some multisets}\\
&\quad\;\; (Y,k),(Z,l) \subseteq (X,m) \text{ s.t. } (X,m) \subseteq (Y,k) \uplus (Z,l).\\
\mathfrak{A} \models_{(X,m)} \forall x \psi \quad &\Leftrightarrow \mathfrak{A} \models_{(X,m)[(A,n)/x]} \psi\\
\mathfrak{A} \models_{(X,m)} \exists x \psi \quad &\Leftrightarrow \mathfrak{A} \models_{(X,m)[F/x]} \psi \text{ holds for some function}\\
&\quad\;\; F: [(X,m)]_{\text{cset}} \rightarrow \mathcal{P}^{+}((A,n)).
\end{aligned}
$$

Multiteam semantics cont.

Definition (Multiteam semantics)

$\mathfrak{A}$ a τ-multistructure, (A, n) the domain of $\mathfrak{A}$, and (X, m) a multiteam over $\mathfrak{A}$.

$$
\begin{array}{lcl}
\mathfrak{A} \models_{(X,m)} x = y & \Leftrightarrow & \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) = s(y) \\
\mathfrak{A} \models_{(X,m)} x \neq y & \Leftrightarrow & \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(x) \neq s(y) \\
\mathfrak{A} \models_{(X,m)} R(\vec{x}) & \Leftrightarrow & \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \in R^{\mathfrak{A}} \\
\mathfrak{A} \models_{(X,m)} \neg R(\vec{x}) & \Leftrightarrow & \forall s \in X: \text{ if } m(s) \geq 1 \text{ then } s(\vec{x}) \notin R^{\mathfrak{A}} \\
\mathfrak{A} \models_{(X,m)} (\psi \wedge \theta) & \Leftrightarrow & \mathfrak{A} \models_{(X,m)} \psi \text{ and } \mathfrak{A} \models_{(X,m)} \theta \\
\mathfrak{A} \models_{(X,m)} (\psi \vee \theta) & \Leftrightarrow & \mathfrak{A} \models_{(Y,k)} \psi \text{ and } \mathfrak{A} \models_{(Z,l)} \theta \text{ for some multisets} \\
 & & (Y,k),(Z,l) \subseteq (X,m) \text{ s.t. } (X,m) \subseteq (Y,k) \uplus (Z,l). \\
\mathfrak{A} \models_{(X,m)} \forall x \psi & \Leftrightarrow & \mathfrak{A} \models_{(X,m)[(A,n)/x]} \psi \\
\mathfrak{A} \models_{(X,m)} \exists x \psi & \Leftrightarrow & \mathfrak{A} \models_{(X,m)[F/x]} \psi \text{ holds for some function} \\
 & & F\colon [(X,m)]_{\mathrm{cset}} \to \mathcal{P}^{+}((A,n)).
\end{array}
$$

Strict multiteam semantics

The so-called *strict multiteam semantics* is obtained from the previous definition by adding the following two requirements.

(i) Disjunction: $(Y, n) \uplus (Z, k) = (X, m)$.

(ii) Existential quantification: for all $s \in X$ and $0 < i \leq m(s)$, $F\big((s, i)\big) = (B, n)$ for some singleton $B = \{b\}$ and $n(b) = 1$.

Multiteam semantics extends team semantics

Proposition

$\mathfrak{A}$ a multistructure with domain (A, n), and (X, m) a multiteam over $\mathfrak{A}$ such that $n(a) = m(s) = 1$ for all $a \in A$ and $s \in X$. Define $\mathfrak{B} := (A, (R^{\mathfrak{A}})_{R \in \tau})$. Then for every $\varphi \in \mathrm{FO}$ it holds that

$$\mathfrak{A} \models_{(X,m)} \varphi \text{ if and only if } \mathfrak{B} \models_X \varphi.$$

Probabilistic inclusion atom

$(X, m)_{\vec{x}=\vec{a}}$ is the multiteam (X, n) where n agrees with m on all assignments $s \in X$ with $s(\vec{x}) = \vec{a}$, and otherwise n maps s to 0.

If $\vec{x}, \vec{y}$ are variable sequences of the same length, then $\vec{x} \leq \vec{y}$ is a *probabilistic inclusion atom* with the following semantics:

$$\mathfrak{A} \models_{(X,m)} \vec{x} \leq \vec{y}$$
$$\text{iff } |(X, m)_{\vec{x}=\vec{s}(\vec{x})}| \leq |(X, m)_{\vec{y}=s(\vec{x})}| \text{ for all } s : \mathrm{Var}(\vec{x}) \to A.$$

Probabilistic inclusion atom

$(X, m)_{\vec{x}=\vec{a}}$ is the multiteam (X, n) where n agrees with m on all assignments $s \in X$ with $s(\vec{x}) = \vec{a}$, and otherwise n maps s to 0.

If $\vec{x}, \vec{y}$ are variable sequences of the same length, then $\vec{x} \leq \vec{y}$ is a *probabilistic inclusion atom* with the following semantics:

$$\mathfrak{A} \models_{(X,m)} \vec{x} \leq \vec{y}$$
$$\text{iff } |(X, m)_{\vec{x}=\vec{s}(\vec{x})}| \leq |(X, m)_{\vec{y}=s(\vec{x})}| \text{ for all } s : \mathrm{Var}(\vec{x}) \to A.$$

Probabilistic interpretation

Multiteams (X, m) induce a natural probability distribution p over the assignments of X. Namely, we define $p\colon X \to [0,1]$ such that

$$p(s) = \frac{m(s)}{\sum_{s\in X} m(s)}.$$

The probability that a tuple of (random) variables $\vec{x}$ takes value $\vec{a}$, written $\Pr(\vec{x} = \vec{a})$, is then

$$\sum_{\substack{s\in X,\\ s(\vec{x})=\vec{a}}} p(s).$$

The probabilistic inclusion atom $\vec{x} \leq \vec{y}$ indicates that $\Pr(\vec{x} = \vec{a}) \leq \Pr(\vec{y} = \vec{a})$ for all values $\vec{a}$. However in the *finite* $\Pr(\vec{x} = \vec{a}) = \Pr(\vec{y} = \vec{a})$ follows.

Probabilistic interpretation

Multiteams (X, m) induce a natural probability distribution p over the assignments of X. Namely, we define $p\colon X \to [0,1]$ such that

$$p(s) = \frac{m(s)}{\sum_{s\in X} m(s)}.$$

The probability that a tuple of (random) variables $\vec{x}$ takes value $\vec{a}$, written $\Pr(\vec{x} = \vec{a})$, is then

$$\sum_{\substack{s\in X,\\ s(\vec{x})=\vec{a}}} p(s).$$

The probabilistic inclusion atom $\vec{x} \leq \vec{y}$ indicates that $\Pr(\vec{x} = \vec{a}) \leq \Pr(\vec{y} = \vec{a})$ for all values $\vec{a}$. However in the *finite* $\Pr(\vec{x} = \vec{a}) = \Pr(\vec{y} = \vec{a})$ follows.

Probabilistic interpretation

Multiteams (X, m) induce a natural probability distribution p over the assignments of X. Namely, we define $p\colon X \to [0,1]$ such that

$$p(s) = \frac{m(s)}{\sum_{s \in X} m(s)}.$$

The probability that a tuple of (random) variables $\vec{x}$ takes value $\vec{a}$, written $\Pr(\vec{x} = \vec{a})$, is then

$$\sum_{\substack{s \in X, \\ s(\vec{x}) = \vec{a}}} p(s).$$

The probabilistic inclusion atom $\vec{x} \leq \vec{y}$ indicates that $\Pr(\vec{x} = \vec{a}) \leq \Pr(\vec{y} = \vec{a})$ for all values $\vec{a}$. However in the *finite* $\Pr(\vec{x} = \vec{a}) = \Pr(\vec{y} = \vec{a})$ follows.

Probabilistic independence

The objective is that that $\mathfrak{A} \models_{(X,m)} \vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ iff for all $\vec{a}\vec{b}\vec{c}$,

$$\Pr(\vec{y} = \vec{b}, \vec{z} = \vec{c} | \vec{x} = \vec{a}) = \Pr(\vec{y} = \vec{b} | \vec{x} = \vec{a}) \Pr(\vec{z} = \vec{c} | \vec{x} = \vec{a}),$$

that is, the probability of $\vec{y} = \vec{b}$ is independent of the probability of $\vec{z} = \vec{c}$, given $\vec{x} = \vec{a}$.

Formally: $\vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ is a *probabilistic conditional independence atom*, defined by

$$\mathfrak{A} \models_{(X,m)} \vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$$

if for all $s \colon \mathrm{Var}(\vec{x}\vec{y}\vec{z}) \to A$ it holds that

$$|(X,m)_{\vec{x}\vec{y}=s(\vec{x}\vec{y})}| \cdot |(X,m)_{\vec{x}\vec{z}=s(\vec{x}\vec{z})}| = |(X,m)_{\vec{x}\vec{y}\vec{z}=s(\vec{x}\vec{y}\vec{z})}| \cdot |(X,m)_{\vec{x}=s(\vec{x})}|.$$

Probabilistic independence

The objective is that that $\mathfrak{A} \models_{(X,m)} \vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ iff for all $\vec{a}\vec{b}\vec{c}$,

$$\Pr(\vec{y} = \vec{b}, \vec{z} = \vec{c} | \vec{x} = \vec{a}) = \Pr(\vec{y} = \vec{b} | \vec{x} = \vec{a}) \Pr(\vec{z} = \vec{c} | \vec{x} = \vec{a}),$$

that is, the probability of $\vec{y} = \vec{b}$ is independent of the probability of $\vec{z} = \vec{c}$, given $\vec{x} = \vec{a}$.

Formally: $\vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ is a *probabilistic conditional independence atom*, defined by

$$\mathfrak{A} \models_{(X,m)} \vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$$

if for all $s\colon \mathrm{Var}(\vec{x}\vec{y}\vec{z}) \to A$ it holds that

$$|(X,m)_{\vec{x}\vec{y}=s(\vec{x}\vec{y})}| \cdot |(X,m)_{\vec{x}\vec{z}=s(\vec{x}\vec{z})}| = |(X,m)_{\vec{x}\vec{y}\vec{z}=s(\vec{x}\vec{y}\vec{z})}| \cdot |(X,m)_{\vec{x}=s(\vec{x})}|.$$

Non-probabilistic atoms

One can also study the usual dependency notions in the multiteam semantics:

Definition

Let $\mathfrak{A}$ be a multistructure, (X, m) a multiteam over $\mathfrak{A}$, and φ of the form $=(\vec{x}, \vec{y})$, $\vec{x} \subseteq \vec{y}$, or $\vec{y} \perp_{\vec{x}} \vec{z}$.

$$\mathfrak{A} \models_{(X,m)} \varphi \text{ iff } \mathfrak{A} \models_{X^+} \varphi,$$

where X^+ is the team $\{s \in X \mid m(s) \geq 1\}$.

Connections between atoms

- Probabilistic independence atom of the form $\vec{x} \perp\!\!\!\perp \vec{x}$ that $\Pr(\vec{x} = \vec{a}) = 1$ for some value $\vec{a}$.
- Probabilistic $\vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{y}$ is equivalent with the non-probabilistic $=\!(\vec{x}, \vec{y})$.
- Marginal indepenence $\vec{x} \perp\!\!\!\perp \vec{x}$ is equivalent with constancy atom $=\!(\vec{x})$.

Connections between atoms cont.

It was shown by Wong 1997 that the generalised multivalued dependency $\vec{x} \multimap\!\!\rightarrow \vec{y}$ holds in an extended relational data model if and only if the underlying relational model satisfies the multivalued dependency $\vec{x} \twoheadrightarrow \vec{y}$. This is stated in the following theorem reformulated into our framework.

Theorem

Let $\mathfrak{A}$ be a multistructure, X a team over $\mathfrak{A}$, and $\vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ a probabilistic conditional independence atom such that $\mathrm{Var}(\vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}) = \mathrm{Dom}(X)$ and $\vec{x}, \vec{y}, \vec{z}$ are pairwise disjoint. Let 1 denote the constant function that maps all assignments of X to 1. Then $\mathfrak{A} \models_{(X,1)} \vec{y} \perp\!\!\!\perp_{\vec{x}} \vec{z}$ iff $\mathfrak{A} \models_{(X,1)} \vec{y} \perp_{\vec{x}} \vec{z}$.

The restriction that $\vec{x}, \vec{y}, \vec{z}$ are disjoint can be now removed.

Locality in multiteams

For $V \subseteq \mathrm{Dom}(X)$, we define $(X, m) \upharpoonright V := (X \upharpoonright V, n)$ where

$$n(s) := \sum_{\substack{s' \in X, \\ s' \upharpoonright V = s}} m(s').$$

The following locality principle holds by easy structural induction.

Proposition (Locality)

Let $\mathfrak{A}$ be a multistructure, (X, m) a multiteam, and V a set of variables such that $\mathrm{Fr}(\varphi) \subseteq V \subseteq \mathrm{Dom}(X)$. Then for all $\varphi \in \mathrm{FO}(\leq, \perp\!\!\!\perp_{\mathrm{c}}, =\!(\cdot), \subseteq, \perp_{\mathrm{c}})$ it holds that $\mathfrak{A} \models_{(X,m)} \varphi$ iff $\mathfrak{A} \models_{(X,m)\upharpoonright V} \varphi$.

Flattness in multiteams

Definition (Weak flatness)

We say that a formula φ is *weakly flat* if for all multistructures $\mathfrak{A}$ and for all multiteams (X, m) it holds that

$$\mathfrak{A} \models_{(X,m)} \varphi \quad \Leftrightarrow \quad \mathfrak{A} \models_{(X,n)} \varphi,$$

where n agrees with m on all s with $m(s) = 0$, and otherwise maps all s to 1. The multiteam (X, n) is then called the *weak flattening* of (X, m). A logic is called *weakly flat* if every formula of this logic is weakly flat.

Dependence, conditional independence, and inclusion atoms are insensitive to multiplicities:

Proposition

$\mathrm{FO}(=(\cdot), \subseteq, \perp_{\mathrm{c}})$ *is weakly flat.*

Union closure in multiteam semantics

A formula φ is union closed (in multiteam setting) if

$$(\mathfrak{A} \models_{(X,m)} \varphi \text{ and } \mathfrak{A} \models_{(Y,n)} \varphi) \Rightarrow \mathfrak{A} \models_{(Z,h)} \varphi, \text{ where } (Z,h) = (X,m) \uplus (Y,n).$$

Proposition

$\mathrm{FO}(\leq, \subseteq)$ *is union closed.*

Strict semantiics

Proposition

Over strict multiteam semantics $\mathrm{FO}(=(\cdot))$ *is weakly flat.*

The logics $\mathrm{FO}(\perp_c)$ and $\mathrm{FO}(\subseteq)$ are not weakly flat under strict multiteam semantics as shown in the next example.
Similarly, one can show that $\mathrm{FO}(\leq, \subseteq)$ is not union closed under strict multiteam semantics. Moreover one can show that locality hold also under strict multiteam semantics.