

Higher-Order Probabilistic Programming

A Tutorial at POPL 2019

Part III

Ugo Dal Lago

(Based on joint work with *Flavien Breuvert*, *Raphaëlle Crubillé*, *Charles Grellois*, *Davide Sangiorgi*, . . .)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



POPL 2019, Lisbon, January 14th

Equivalence and Distance Checking

- ▶ How could we check two higher-order probabilistic programs to be (context) equivalent?

Equivalence and Distance Checking

- ▶ How could we check two higher-order probabilistic programs to be (context) equivalent?
- ▶ How about their distance?

Equivalence and Distance Checking

- ▶ How could we check two higher-order probabilistic programs to be (context) equivalent?
- ▶ How about their distance?
- ▶ Contextual equivalence and contextual distance are good answers, *definitionally*.
 - ▶ They are the coarsest compatible and adequate relation and metric between programs.
 - ▶ There is however an explicit quantification over all contexts, which make argument inherently complicated.

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

Examples

$\lambda x.x$

$I \oplus \Omega$

vs.

I

Examples

$$\Delta\Delta = (\lambda x.xx)(\lambda x.xx)$$

$$I \oplus \Omega \quad \text{vs.} \quad I$$

Examples

Not Context Equivalent: $C = [\cdot]$.

Context Distance? Consider $C_n = (\lambda x. \underbrace{x \dots x}_{n \text{ times}})[\cdot]$.

$I \oplus \Omega$ vs. I

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

$$I \oplus \Omega \quad \text{vs.} \quad \Omega$$

Examples

Not Context Equivalent: $C = [\cdot]$.

Context Distance? Cannot Easily Amplify.

$$I \oplus \Omega \quad \text{vs.} \quad I$$

$$I \oplus \Omega \quad \text{vs.} \quad \Omega$$

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

$$I \oplus \Omega \quad \text{vs.} \quad \Omega$$

$$(\lambda x. I) \oplus (\lambda x. \Omega) \quad \text{vs.} \quad \lambda x. I \oplus \Omega$$

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

Not Context Equivalent in CBV: $C = (\lambda x.x(xI))[\cdot]$
Apparently Context Equivalent in CBN.

$$I \oplus \Omega \quad \text{vs.} \quad \Omega$$

$$(\lambda x.I) \oplus (\lambda x.\Omega) \quad \text{vs.} \quad \lambda x.I \oplus \Omega$$

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

$$I \oplus \Omega \quad \text{vs.} \quad \Omega$$

$$(\lambda x. I) \oplus (\lambda x. \Omega) \quad \text{vs.} \quad \lambda x. I \oplus \Omega$$

$$Y_1 \quad \text{vs.} \quad Y_2$$

Examples

$$I \oplus \Omega \quad \text{vs.} \quad I$$

$$Y_1 M \rightarrow^* M(Y_2 M) \oplus M(Y_3 M)$$

$$Y_2 M \rightarrow^* M(Y_1 M) \oplus M(Y_3 M)$$

$$Y_3 M \rightarrow^* M(Y_1 M) \oplus M(Y_2 M)$$

$$Y_1 \quad \text{vs.} \quad Y_2$$

Probabilistic Bisimulation in the Abstract [LS1992]

- ▶ **Labelled Markov Chain (LMC)**: a triple $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$, where
 - ▶ \mathcal{S} is a countable set of *states*;
 - ▶ \mathcal{L} is a set of *labels*;
 - ▶ \mathcal{P} is a *transition probability matrix*, i.e., a function $\mathcal{P} : \mathcal{S} \times \mathcal{L} \times \mathcal{S} \rightarrow \mathbb{R}$ such that for every state t and for every label ℓ , $\mathcal{P}(t, \ell, \mathcal{S}) = \sum_{s \in \mathcal{S}} \mathcal{P}(t, \ell, s) \leq 1$;
- ▶ **Bisimulation**: equivalence relation \mathcal{R} on \mathcal{S} such that whenever $t \mathcal{R} s$, it holds that $\mathcal{P}(t, \ell, E) = \mathcal{P}(s, \ell, E)$ for every equivalence class E of \mathcal{S} modulo \mathcal{R} .
- ▶ Variation: **Simulation**, which is required to be a preorder.
- ▶ Bisimilarity and Similarity can always be formed.

Proposition

$$\sim = \preceq \cap \preceq^{op}.$$

A Labelled Markov Chain for Λ_{\oplus}

Terms

A Labelled Markov Chain for Λ_{\oplus}

Terms

Values

A Labelled Markov Chain for Λ_{\oplus}

Terms

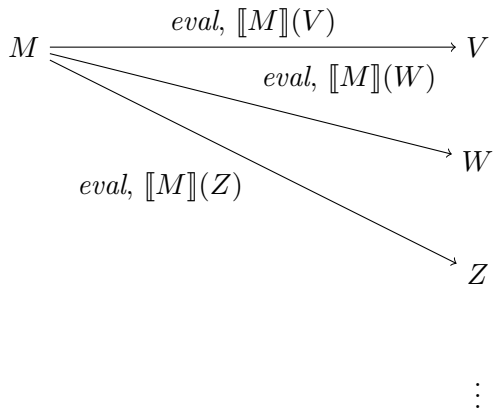
Values

M

A Labelled Markov Chain for Λ_{\oplus}

Terms

Values



A Labelled Markov Chain for Λ_{\oplus}

Terms

Values

$\lambda x.N$

A Labelled Markov Chain for Λ_{\oplus}

Terms

Values

$$N\{W/x\} \xleftarrow{W, 1} \lambda x.N$$

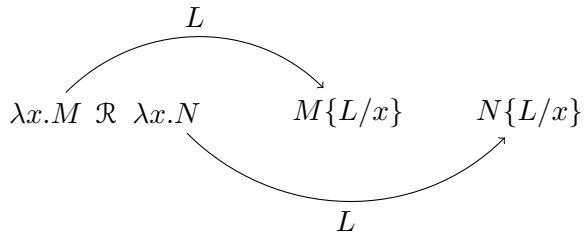
Probabilistic Applicative Bisimulation

$$\lambda x.M \mathcal{R} \lambda x.N$$

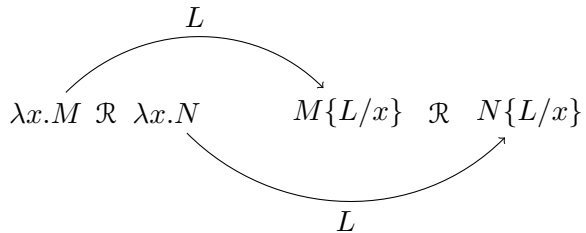
Probabilistic Applicative Bisimulation

$$\lambda x.M \mathcal{R} \lambda x.N \quad \begin{array}{c} L \\ \curvearrowright \\ M\{L/x\} \end{array}$$

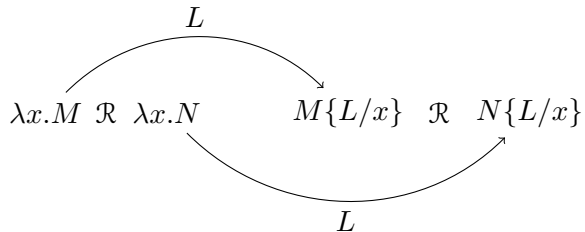
Probabilistic Applicative Bisimulation



Probabilistic Applicative Bisimulation

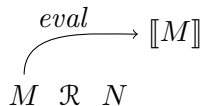
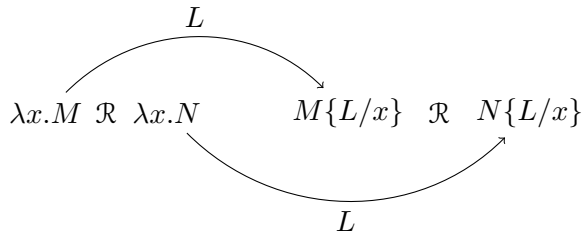


Probabilistic Applicative Bisimulation

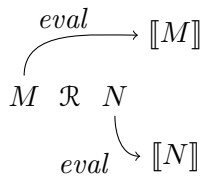
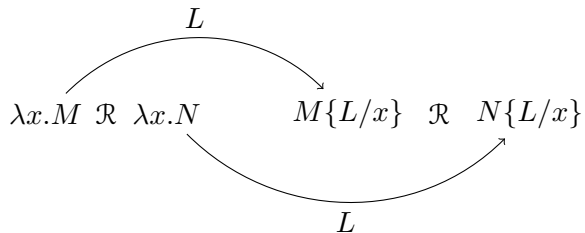


$$M \ \mathcal{R} \ N$$

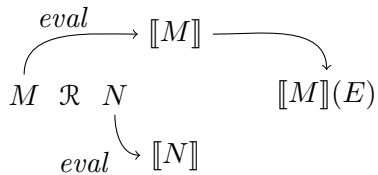
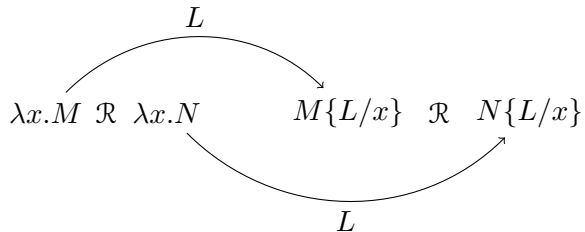
Probabilistic Applicative Bisimulation



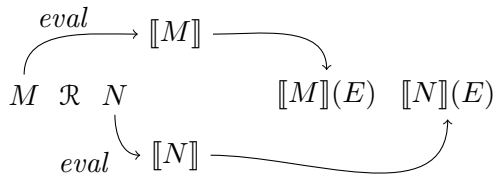
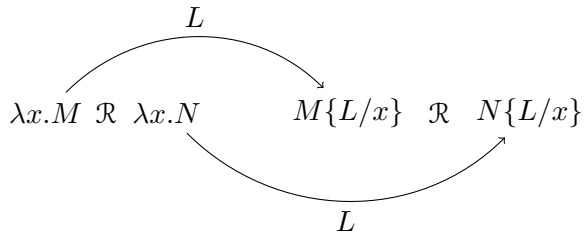
Probabilistic Applicative Bisimulation



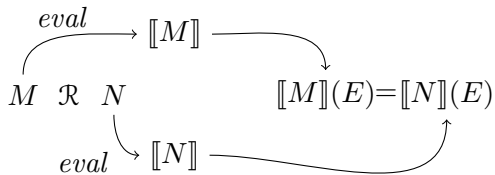
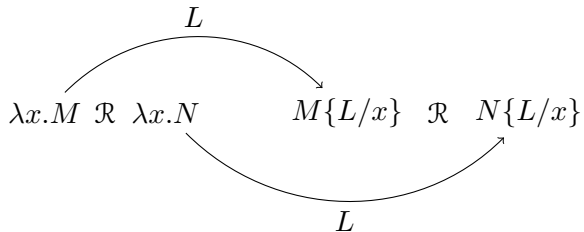
Probabilistic Applicative Bisimulation



Probabilistic Applicative Bisimulation



Probabilistic Applicative Bisimulation



Applicative Bisimilarity vs. Context Equivalence

- ▶ **Bisimilarity**: the union \sim of all bisimulation relations.
- ▶ Is it that \sim is included in \equiv ? How to prove it?
- ▶ Natural strategy: is \sim a congruence?
 - ▶ If this is the case:

$$\begin{aligned} M \sim N &\implies C[M] \sim C[N] \implies \sum \llbracket C[M] \rrbracket = \sum \llbracket C[N] \rrbracket \\ &\implies M \equiv N. \end{aligned}$$

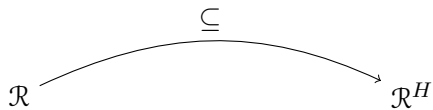
- ▶ This is a necessary sanity check anyway.
- ▶ The naïve proof by induction **fails**, due to application: from $M \sim N$, one cannot directly conclude that $LM \sim LN$.

Howe's Technique

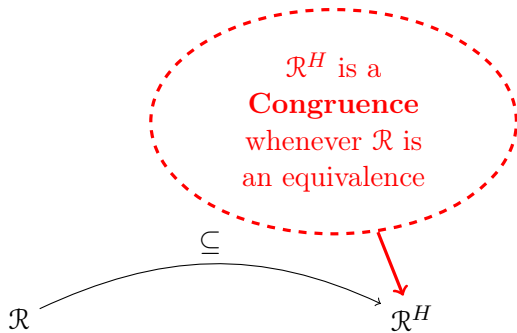
\mathcal{R}

\mathcal{R}^H

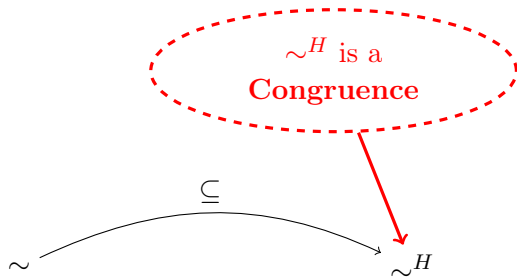
Howe's Technique



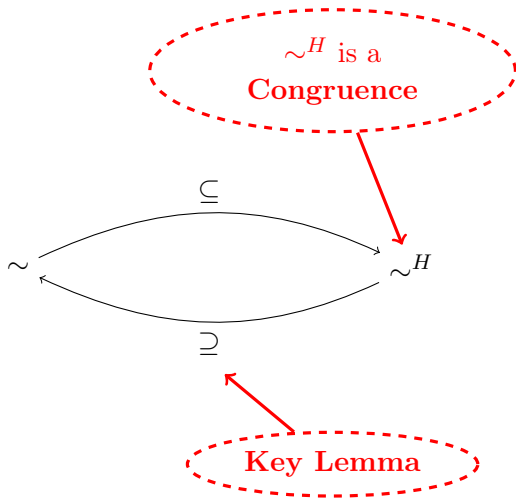
Howe's Technique



Howe's Technique



Howe's Technique



Our Neighborhood

- Λ , where we observe **convergence**

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✓
<i>CBV</i>	✓	✓

[Abramsky1990, Howe1993]

- Λ_{\oplus} with nondeterministic semantics, where we observe **convergence**, in its **may** or **must** flavors.

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✗

[Ong1993, Lassen1998]

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✓

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✓

- ▶ Counterexample for CBN: $(\lambda x.I) \oplus (\lambda x.\Omega) \not\sim \lambda x.I \oplus \Omega$
- ▶ **Where** these discrepancies come from?

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✓

- ▶ Counterexample for CBN: $(\lambda x.I) \oplus (\lambda x.\Omega) \not\sim \lambda x.I \oplus \Omega$
- ▶ **Where** these discrepancies come from?
- ▶ From **testing**!

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\sim \subseteq \equiv$	$\equiv \subseteq \sim$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✓

- ▶ Counterexample for CBN: $(\lambda x.I) \oplus (\lambda x.\Omega) \not\sim \lambda x.I \oplus \Omega$
- ▶ **Where** these discrepancies come from?
- ▶ From **testing**!
- ▶ Bisimulation can be characterized by testing equivalence as follows:

Calculus	Testing
Λ	$T ::= \omega \mid a \cdot T$
probabilistic Λ_{\oplus}	$T ::= \omega \mid a \cdot T \mid \langle T, T \rangle$
nondeterministic Λ_{\oplus}	$T ::= \omega \mid a \cdot T \mid \bigwedge_{i \in I} T_i \mid \dots$

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\preceq \subseteq \leq$	$\leq \subseteq \preceq$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✗

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\preceq \subseteq \leq$	$\leq \subseteq \preceq$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✗

- ▶ Probabilistic simulation can be characterized by testing as follows:

$$T ::= \omega \mid a \cdot T \mid \langle T, T \rangle \mid T \vee T$$

The Probabilistic Case

- ▶ Λ_{\oplus} with probabilistic semantics.

	$\preceq \subseteq \leq$	$\leq \subseteq \preceq$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✗

- ▶ Probabilistic simulation can be characterized by testing as follows:

$$T ::= \omega \mid a \cdot T \mid \langle T, T \rangle \mid T \vee T$$

- ▶ Full abstraction can be recovered if endowing Λ_{\oplus} with parallel disjunction [CDLSV2015].

	$\preceq \subseteq \leq$	$\leq \subseteq \preceq$
<i>CBN</i>	✓	✗
<i>CBV</i>	✓	✓

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega;$

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.

Context Distance: the Affine Case [CDL2015]

$$\frac{}{\Gamma, x \vdash x} \quad \frac{x, \Gamma \vdash M}{\Gamma \vdash \lambda x.M} \quad \frac{\Gamma \vdash M \quad \Delta \vdash N}{\Gamma, \Delta \vdash MN} \quad \frac{\Gamma \vdash M \quad \Gamma \vdash N}{\Gamma \vdash M \oplus N}$$

- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
- ▶ **Behavioural Distance** δ^b .
 - ▶ The metric analogue to bisimilarity.

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
- ▶ **Behavioural Distance** δ^b .
 - ▶ The metric analogue to bisimilarity.
- ▶ **Trace Distance** δ^t .
 - ▶ The maximum distance induced by traces, i.e., sequences of actions:
$$\delta^t(M, N) = \sup_{\mathsf{T}} |Pr(M, \mathsf{T}) - Pr(N, \mathsf{T})|.$$

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
- ▶ **Behavioural Distance** δ^b .
 - ▶ The metric analogue to bisimilarity.
- ▶ **Trace Distance** δ^t .
 - ▶ The maximum distance induced by traces, i.e., sequences of actions:
$$\delta^t(M, N) = \sup_{\mathsf{T}} |Pr(M, \mathsf{T}) - Pr(N, \mathsf{T})|.$$
- ▶ **Soundness and Completeness Results:**

$\delta^b \leq \delta^c$	$\delta^c \leq \delta^b$	$\delta^t \leq \delta^c$	$\delta^c \leq \delta^t$
✓	✗	✓	✓

Context Distance: the Affine Case [CDL2015]

- ▶ Let us consider a simple fragment of Λ_{\oplus} , first.
- ▶ **Preterms:** $M, N ::= x \mid \lambda x.M \mid MM \mid M \oplus M \mid \Omega$;
- ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
- ▶ **Behavioural Distance** δ^b .
 - ▶ The metric analogue to bisimilarity.
- ▶ **Trace Distance** δ^t .
 - ▶ The maximum distance induced by traces, i.e., sequences of actions:
$$\delta^t(M, N) = \sup_{\mathsf{T}} |Pr(M, \mathsf{T}) - Pr(N, \mathsf{T})|.$$
- ▶ **Soundness and Completeness Results:**

$\delta^b \leq \delta^c$	$\delta^c \leq \delta^b$	$\delta^t \leq \delta^c$	$\delta^c \leq \delta^t$
✓	✗	✓	✓

- ▶ **Example:** $\delta^t(I, I \oplus \Omega) = \delta^t(I \oplus \Omega, \Omega) = \frac{1}{2}$.

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.
- ▶ The underlying LMC **does not** reflect copying.

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.
- ▶ The underlying LMC **does not** reflect copying.
- ▶ **A Tuple LMC.**
 - ▶ **Preterms:** $M ::= x \mid \lambda x.M \mid \lambda!x.M \mid MM \mid M \oplus M \mid !M$
 - ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
 - ▶ **States:** *sequences* of terms, rather than terms.
 - ▶ **Actions** not only model parameter passing, but also *copying* of terms.

Context

$\frac{}{!\Gamma, x \vdash x}$	$\frac{}{!\Gamma, !x \vdash x}$	$\frac{x, \Gamma \vdash M}{\Gamma \vdash \lambda x.M}$	$\frac{!x, \Gamma \vdash M}{\Gamma \vdash \lambda !x.M}$
$\frac{!\Gamma \vdash M}{!\Gamma \vdash !M}$	$\frac{\Gamma, !\Theta \vdash M \quad \Delta, !\Theta \vdash N}{\Gamma, \Delta, \Theta \vdash MN}$	$\frac{\Gamma \vdash M \quad \Gamma \vdash N}{\Gamma \vdash M \oplus N}$	

► N

► T

► A Tuple LMC.

- **Preterms:** $M ::= x \mid \lambda x.M \mid \lambda !x.M \mid MM \mid M \oplus M \mid !M$
- **Terms:** any preterm M such that $\Gamma \vdash M$.
- **States:** *sequences* of terms, rather than terms.
- **Actions** not only model parameter passing, but also *copying* of terms.

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.
- ▶ The underlying LMC **does not** reflect copying.
- ▶ **A Tuple LMC.**
 - ▶ **Preterms:** $M ::= x \mid \lambda x.M \mid \lambda!x.M \mid MM \mid M \oplus M \mid !M$
 - ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
 - ▶ **States:** *sequences* of terms, rather than terms.
 - ▶ **Actions** not only model parameter passing, but also *copying* of terms.
- ▶ **Soundness and Completeness Results:**

$\delta^t \leq \delta^c$	$\delta^c \leq \delta^t$
✓	✓

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.
- ▶ The underlying LMC **does not** reflect copying.
- ▶ **A Tuple LMC.**
 - ▶ **Preterms:** $M ::= x \mid \lambda x.M \mid \lambda !x.M \mid MM \mid M \oplus M \mid !M$
 - ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
 - ▶ **States:** *sequences* of terms, rather than terms.
 - ▶ **Actions** not only model parameter passing, but also *copying* of terms.
- ▶ **Soundness and Completeness Results:**

$\delta^t \leq \delta^c$	$\delta^c \leq \delta^t$
✓	✓

- ▶ **Examples:** $\delta^t(! (I \oplus \Omega), !\Omega) = \frac{1}{2}$ $\delta^t(! (I \oplus \Omega), !I) = 1$.

Context Distance: the General Case [CDL2016]

- ▶ None of the abstract notions of distance δ gives us that $\delta(I, I \oplus \Omega) = 1$.
- ▶ The underlying LMC **does not** reflect copying.
- ▶ **A Tuple LMC.**
 - ▶ **Preterms:** $M ::= x \mid \lambda x.M \mid \lambda !x.M \mid MM \mid M \oplus M \mid !M$
 - ▶ **Terms:** any preterm M such that $\Gamma \vdash M$.
 - ▶ **States:** *sequences* of terms, rather than terms.
 - ▶ **Actions** not only model parameter passing, but also *copying* of terms.
- ▶ **Soundness and Completeness Results:**

$\delta^t \leq \delta^c$	$\delta^c \leq \delta^t$
✓	✓

- ▶ **Examples:** $\delta^t(! (I \oplus \Omega), !\Omega) = \frac{1}{2}$ $\delta^t(! (I \oplus \Omega), !I) = 1$.
- ▶ **Trivialisation** does not hold in general, but becomes true in *strongly normalising* fragments or in presence of *parallel disjunction*.

Denotational Models

- ▶ **Probabilistic Powerdomains** [JonesPlotkin1991, JungTix1998].
 - ▶ Probabilistic effects are interpreted in *monadic* style [Moggi1989].
 - ▶ Inherently difficult endeavour, because the category of measurable spaces is not cartesian closed.

Denotational Models

- ▶ **Probabilistic Powerdomains** [JonesPlotkin1991, JungTix1998].
 - ▶ Probabilistic effects are interpreted in *monadic* style [Moggi1989].
 - ▶ Inherently difficult endeavour, because the category of measurable spaces is not cartesian closed.
- ▶ **Probabilistic Coherence Spaces** [DanosEhrhard2011, EPT2016].
 - ▶ The difficulties with powerdomains are circumvented through Girard's quantitative semantics.
 - ▶ Proved fully abstract for discrete probabilistic choice.
 - ▶ Can be adapted to deal with continuous distributions [EPT2018].

Denotational Models

- ▶ **Probabilistic Powerdomains** [JonesPlotkin1991, JungTix1998].
 - ▶ Probabilistic effects are interpreted in *monadic* style [Moggi1989].
 - ▶ Inherently difficult endeavour, because the category of measurable spaces is not cartesian closed.
- ▶ **Probabilistic Coherence Spaces** [DanosEhrhard2011, EPT2016].
 - ▶ The difficulties with powerdomains are circumvented through Girard's quantitative semantics.
 - ▶ Proved fully abstract for discrete probabilistic choice.
 - ▶ Can be adapted to deal with continuous distributions [EPT2018].
- ▶ **Quasi-Borel Spaces** [HKSY2017, VKS2019].
 - ▶ Another very clever way to side-step the problems one faces in the category of measurable spaces.
 - ▶ Can adequately model continuous distributions *and* conditioning.

Denotational Models

- ▶ **Probabilistic Powerdomains** [JonesPlotkin1991, JungTix1998].
 - ▶ Probabilistic effects are interpreted in *monadic* style [Moggi1989].
 - ▶ Inherently difficult endeavour, because the category of measurable spaces is not cartesian closed.
- ▶ **Probabilistic Coherence Spaces** [DanosEhrhard2011, EPT2016].
 - ▶ The difficulties with powerdomains are circumvented through Girard's quantitative semantics.
 - ▶ Proved fully abstract for discrete probabilistic choice.
 - ▶ Can be adapted to deal with continuous distributions [EPT2018].
- ▶ **Quasi-Borel Spaces** [HKSY2017, VKS2019].
 - ▶ Another very clever way to side-step the problems one faces in the category of measurable spaces.
 - ▶ Can adequately model continuous distributions *and* conditioning.
- ▶ **Game and GoI Models** [DanosHarmer2002, DLFVY2017].
 - ▶ Higher-order programs are interpreted as strategies or automata.
 - ▶ Game models are fully abstract, in presence of states.

Thank You!

Questions?