Context Distances for Probabilistic Higher-Order Languages

Raphaëlle Crubillé
joint work with Houssein Mansour

LIS, Aix-Marseille Université

October 24, 2024

Equivalence for Higher-Order Probabilistic Programs

Randomized higher-order languages

$$M := x \mid \lambda x.M \mid (M)N \mid \ldots \mid M \oplus_p N.$$

Definition (Morris contextual equivalence)

$M \equiv^{ctx} N$ iff $\forall$ C context, $\mathrm{Obs}(C[M]) = \mathrm{Obs}(C[N])$.

Contexts and Observation

- **Context**: a program with a hole $[\cdot]$.
  Model the external environment.
- A choice of **observable**,e.g.:
  - $\mathrm{Proba}(M \downarrow)$
  - $\mathrm{Proba}(M \downarrow 0)$ $\Rightarrow$ only ground-type programs can be observed.

Semantics as an Interactive System

Pure λ-Terms: Applicative Bisimulation

LTS: labels:= Λ-terms, states:= Λ-terms $\sqcup$ Values.

Terms **Values**

$$M \xrightarrow{eval} \lambda x.L$$

$$N\{L/x\} \xleftarrow{L} \lambda x.N$$

Theorem (Abramsky93)

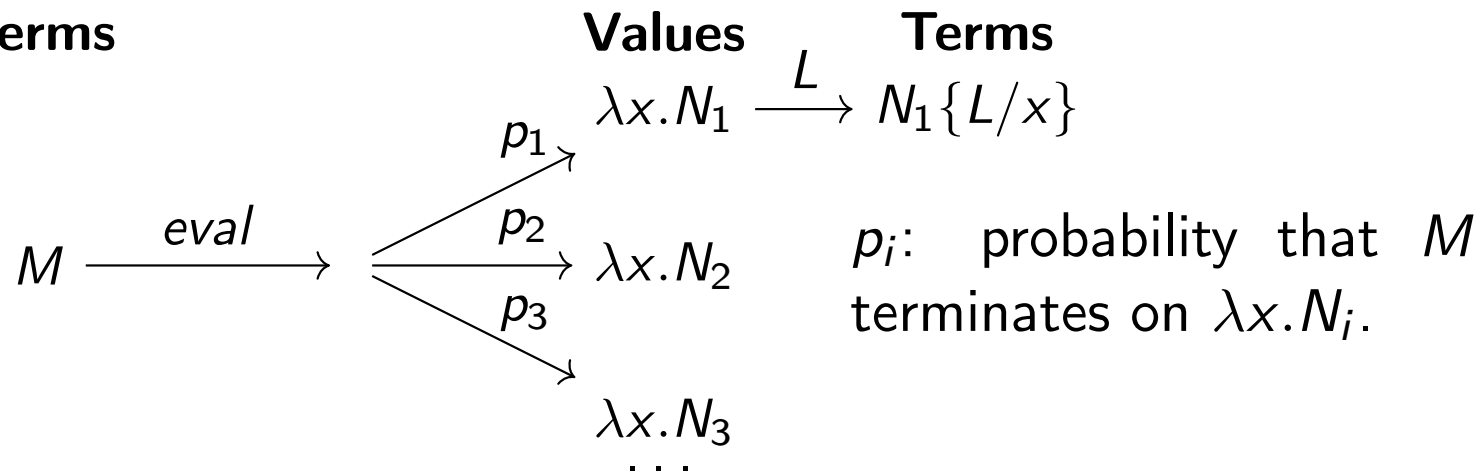
$\forall M, N$ *pure* λ*-terms:* $M \equiv^{ctx} N$ *iff* M, N *are bisimilar.*

Idea

- semantics seen as interaction program - environment
- Use tools developped on transition systems to define equivalence.

Probabilistic Semantics as an Interactive System

pLTS for a probabilistic λ-calculus



Theorem

$\forall M, N$ *pure probabilistic* λ*-terms:*

- *call-by-name evaluation paradigm:* $M \equiv^{ctx} N$ *iff* M, N *are trace equivalent. [Alberti, Dal Lago,Sangiorgi'14]*
- *call-by-value evaluation paradigm:* $M \equiv^{ctx} N$ *iff* M, N *are bisimilar. [C, Dal Lago'14]*

Quantitative generalisation of contextual equivalence ?

Example

M and N are not contextually equivalent, but behave **closely**.

$$M = \lambda f.f(0)$$
$$N = \lambda f.(f(0) \oplus^{\epsilon} f(1)) \qquad \text{où } \epsilon \ll 1.$$

Problem:

How to generalise contextual equivalence in order to compare **quantitatively** two programs ?

A (naive) Definition for Contextual Distance

Definition (Contextual Distance)

$$\delta^{\mathrm{ctx}}(M, N) = \sup_{\mathcal{C} \text{ a context,}} |Obs(C[M]) - Obs(C[N])|$$

A (naive) Definition for Contextual Distance

Definition (Contextual Distance)

$$\delta^{\mathrm{ctx}}(M,N) = \sup_{\mathcal{C} \text{ a context,}} |Prob(\mathcal{C}[M] \downarrow 0) - Prob(\mathcal{C}[N] \downarrow 0)|$$

A (naive) Definition for Contextual Distance

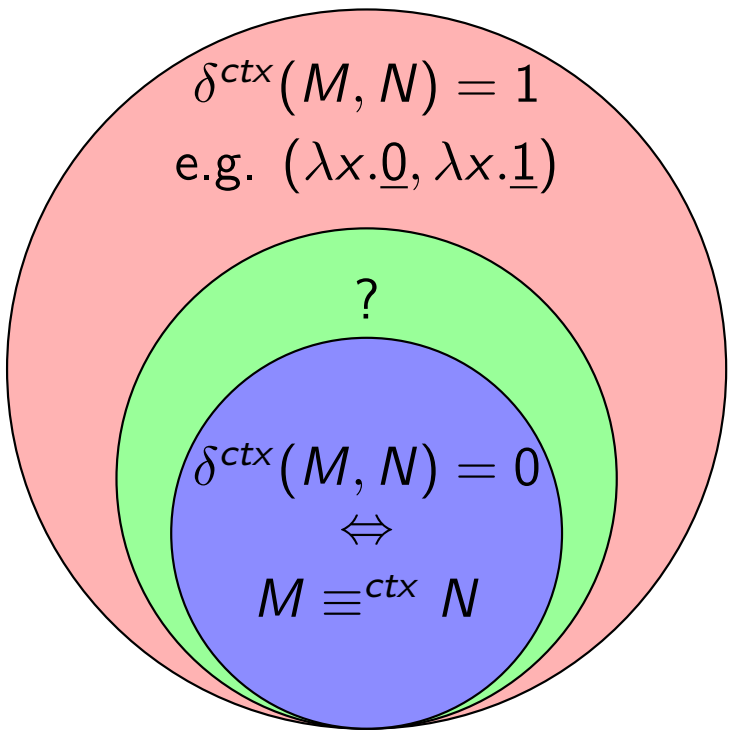
Definition (Contextual Distance)

$$\delta^{\mathrm{ctx}}(M,N) = \sup_{\mathcal{C} \text{ a context,}} |Prob(\mathcal{C}[M] \downarrow 0) - Prob(\mathcal{C}[N] \downarrow 0)|$$

The Trivialization Problem

Is contextual distance really **finer** than contextual equivalence ?

i.e. does there exist M, N such that $0 < \delta^{\mathrm{ctx}}(M,N) < 1$?

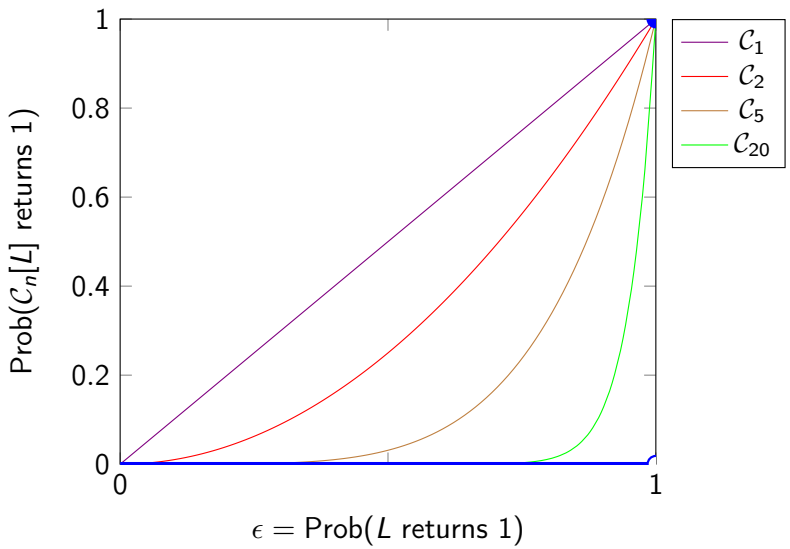


Trivialization Phenomenon: an Example

Two programs (intuitively) close:

$$M = \underline{1} \qquad N_\epsilon = (\underline{1} \oplus^\epsilon \underline{0}) \qquad \text{with } \epsilon \ll 1$$

Computing $\delta^{ctx}(M, N)$:



Each context $\mathcal{C}_n$:

- **copies** n times its argument
- evaluates all copies
- returns 1 if all instances return $1, 0$ otherwise.

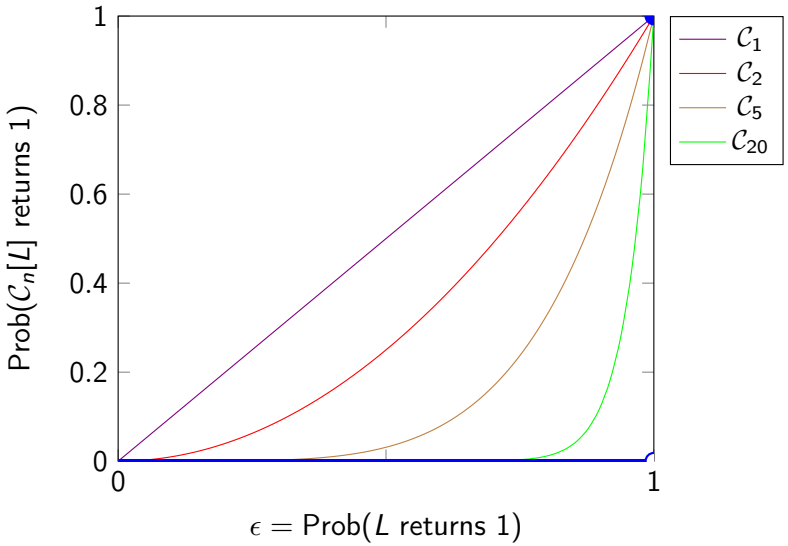
$$\forall \epsilon > 0,\ \delta^{ctx}(M, N_\epsilon) = 1.$$

Trivialization Phenomenon: an Example

Two programs (intuitively) close:

$$M = \underline{1} \qquad N_\epsilon = (\underline{1} \oplus^\epsilon \underline{0}) \qquad \text{with } \epsilon \ll 1$$

Computing $\delta^{ctx}(M, N)$:



Each context $\mathcal{C}_n$:

- **copies** n times its argument
- evaluates all copies
- returns 1 if all instances return $1, 0$ otherwise.

$$\forall \epsilon > 0,\ \delta^{ctx}(M, N_\epsilon) = 1.$$

Question: Can we separate similarly any observable programs M, N with $Obs(M) \neq Obs(N)$. ?

Results on trivialisation

Theorem (C., Dal Lago 2017)

In the $\mathbb{T}_{\oplus}$ language – where all programs terminate with probability 1, contextual distance trivialises.

Theorem (C., Dal Lago 2017)

The contextual distance trivialize in the language $\Lambda^{\vee}_{\oplus}$ where:

- *there are non-terminating programs,*
- *contexts **expressive power** is extended by Abramsky's parallel convergence testers,*

Proofs

- the original ones: based on techniques from real analysis (Bernstein polynomials, dominated convergence theorem).
- new, more "intrinsically probabilistic" proof: based on the law of large numbers [C.,Mansour'24]

The (weak) law of large numbers

Law of Large Numbers [Bernouilli'1713]

(X_k) a sequence of independant identically distributed random variables. $\forall \epsilon > 0$:

$$\lim_{k \to \infty} \text{Proba}(|\frac{1}{k} \sum_{i=1}^{k} X_i - \mathbb{E}(X_i)| \leq \epsilon) = 1.$$

Testing contexts in the $T_{\oplus}$ language

Suppose M is of type Bool: $Obs(M) = Proba(M \downarrow 1)$.
We build a family of testing contexts

Construction of the k-th testing context

Let fix $E, \epsilon \in \mathbb{Q} \cap [0, 1]$.
We build a context $C_n^{E,\epsilon}$ that:

- takes n copies of its argument;
- evaluates each copy x_i independently, and computes $b_i := (x_i? = 0)$
- compute $s = \sum_{i=1\ldots k} b_i$, and test whether $|\frac{s}{k} - E| > \epsilon$.

From testing contexts to amplification contexts.

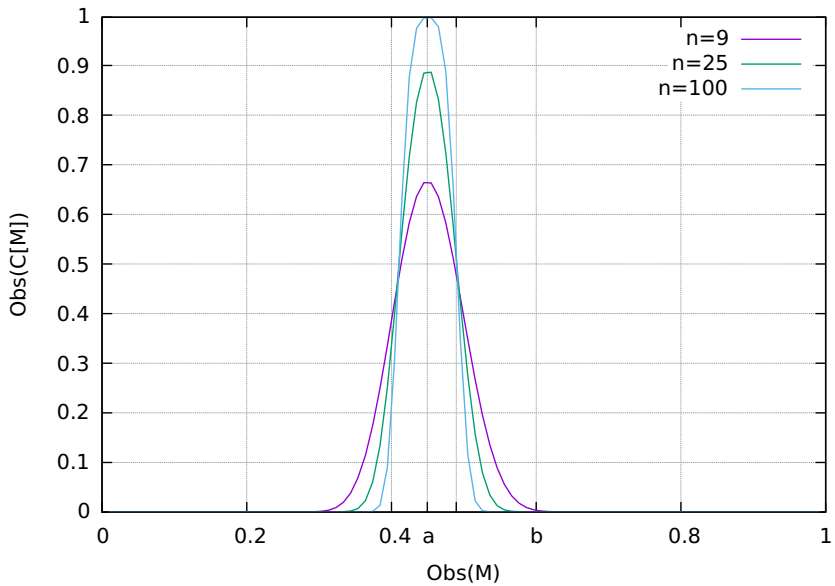
Applying the LLN to testing contexts:

$$\lim_{k\to\infty} \mathrm{Obs}(C_k^{E,\epsilon}[M]) = \begin{cases} 1 \text{ when } \mathrm{Obs}(M) \in (E-\epsilon, E+\epsilon) \\ 0 \text{ when } \mathrm{Obs}(M) \notin [E-\epsilon, E+\epsilon]. \end{cases}$$

Proof: $Obs(C_k^{E,\epsilon}[M]) = Proba(|\frac{1}{k}\sum_{i=1}^{k} X_i^M - E| \leq \epsilon))$
We use LLN + $(\mathbb{E}(X_i^M) = \mathrm{Obs}(M))$.

Illustration:

Asymptotic approximation of the behaviour of $C_n^{E,\epsilon}$,
$E = a = 0,45$, $\epsilon = 0,04$
(with the central limit theorem).



Testing contexts in $\Lambda_{\oplus}^{\vee_{n,m}}$

$\vee_{n,m}$: n-ary parallel convergence tester, $\forall n, m \in \mathbb{N}$.
$\vee_{n,m}(M_1, \ldots, M_n)$ terminates iff at least m of the n M_i terminates.

Observable programs in $\Lambda_{\oplus}^{\vee}$.

- Observation is done at all types
- $\mathrm{Obs}(M) = Proba(M\downarrow)$.

Testing contexts:

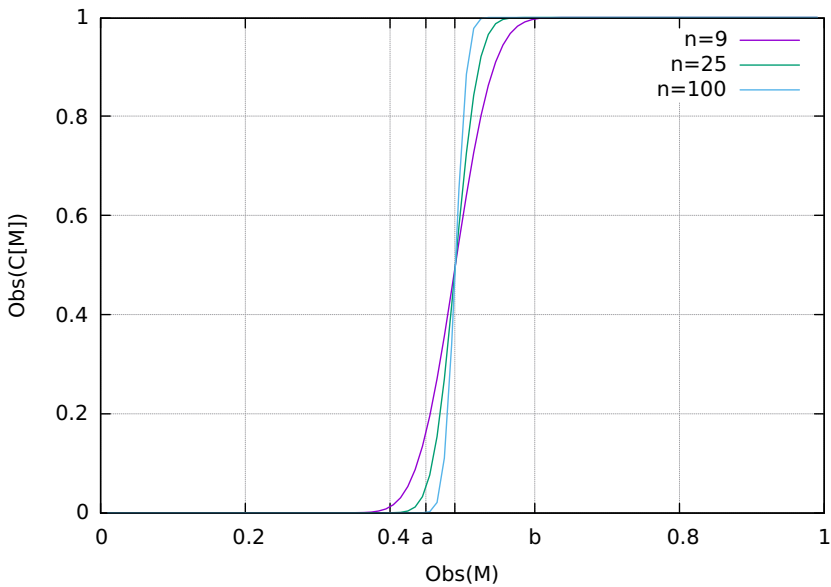
$Obs(C_k^{E,\epsilon}[M]) =$
$Proba(\frac{1}{k}\sum_{i=1}^{k} X_i^M \geq E + \epsilon))$

Construction of the n-th testing context

Let fix $E = \frac{E_1}{E_2}, \epsilon = \frac{\epsilon_1}{\epsilon_2} \in \mathbb{Q}$.

We build a context $C_n^{E,\epsilon}$ that:

- takes n copies $x_1, \ldots, x_n$ of $[\cdot]$
- m: smallest integer s.t. $\frac{m}{n} \geq E + \epsilon$.
- computes $\vee_{n,m}(x, \ldots, x)$



$\Lambda^{\vee}_{\oplus} = \Lambda_{\oplus}$ + basic 2-ary convergence tester ?

We can simulate in this language:

- $\vee_{n,1}(M_1, \ldots, M_n)$: stops iff only 1 of the n copies terminate
- $\vee_{n,n}(M_1, \ldots, M_n)$: stops iff only n of the n copies terminate
- but not the others $\vee_{n,m}$ $\quad \Rightarrow$ contrary to the deterministic case.

Conclusion

- It seems difficult to use the LLN to show trivialization in $\Lambda^{\vee}_{\oplus}$.
- Nonetheless, can we find a more elegant proof than the one in [CDL'17] ?

$\Lambda_{\oplus}^{\vee} = \Lambda_{\oplus}$ + basic 2-ary convergence tester ?

We can simulate in this language:

- $\vee_{n,1}(M_1, \ldots, M_n)$: stops iff only 1 of the n copies terminate
- $\vee_{n,n}(M_1, \ldots, M_n)$: stops iff only n of the n copies terminate
- but not the others $\vee_{n,m}$ $\quad \Rightarrow$ contrary to the deterministic case.

Conclusion

- It seems difficult to use the LLN to show trivialization in $\Lambda_{\oplus}^{\vee}$.
- Nonetheless, can we find a more elegant proof than the one in [CDL'17] ?
- Yes, using Threshold theory .

Threshold Phenomena Theory

Threshold phenomenon: when the probability for an event changes sharply for a slight change of an underlying parameter.

Success probability of f: $\mathbb{B}^n \to \mathbb{B}$ with random input:

$$\begin{aligned}\hat{f} :&[0,1] \to [0,1]\\ &x \mapsto \mathit{Proba}[f(X_1,\ldots,X_n) = 1 \mid X_i \leftarrow \text{Bernouilli}(x)]\end{aligned}$$

The Threshold Question:

When f is non-constant and monotonous:
What it the **width** of the window in which $\hat{f}$ goes from "almost 0" to "almost 1" ?

- **Applications:** Random graphs [Erdös-Renyi'60], Physics (Phase transitions, Critical percolation), Voting Systems...
- **Survey**: *Sharp threshold phenomena in statistical physics*, [Duminil-Copin'18]

The Tribes Functions

Example ($Tribes : \mathbb{B}^{n.m} \to \mathbb{B}$)

- the variables are grouped in n tribes of m individuals;
- $Tribes(x_1, \ldots, x_{n \cdot m})$ is 1 when in each tribe, there is at least 1 variable with value 1.

Properties

- Tribes is known to have a sharp threeshold (because it is **weakly symmetric**)
- The critical probability can be choosed while fixing m and n.
- Tribes can be simulated by a context in $\Lambda^{\vee}_{oplus}$:
  $C[x_1, \ldots, x_{nm}] :=$
  $(\lambda y_1 \lambda y_2 \ldots \lambda y_n . y_n) \vee_{1,n}(x_1, \ldots, x_n) \ldots \vee_{1,n}(x_{(n-1)m+1}, \ldots, x_{nm})$.

Alternative Definitions for Context Distances

Open(?) Question: What kind of contexts do we want to consider for defining contextual distance ?

Linear (or affine) contexts

- contexts can uses their argument (at most) one time [CDL'15]
- contextual distance is caracterised by trace distance on the applicative LMC.

Example

$M := \text{ let } x \leftarrow \text{random}\{1, \ldots, n\} \text{ in } (\lambda y.(\text{ if } x = y \text{ then } 1 \text{ else } 0) \oplus_{\frac{1}{2}} x)$

$N := \text{ let } x \leftarrow \text{random}\{1, \ldots, n\} \text{ in } (\lambda y.0 \oplus_{\frac{1}{2}} x)$

- for linear contexts: $\delta(M, N) = \frac{1}{n}$
- for contexts that can use their arguments *twice*: $\delta(M, N) \geq \frac{1}{4}$

The increasing rate of $\delta(M, N)$ is **faster** than the one caused by the generic amplification phenomenon.

Work in progress

Bounded Contextual Distances

$$\delta^n(M,N) = \sup_{\substack{\text{C context using at} \\ \text{most n times its} \\ \text{argument}}} |Obs(C[M]) - Obs(C[N])|$$

Tool: type system for contexts based on Linear Logic.
Conjecture (?): Bounded contextual distances could be caracterised as graded trace distance on a LMC.

Lower Bounds from our testing contexts:

- We can obtain a lower bound on the growing speed of the $\delta^n(M,N)$.
- Central Limit Theorem (CLT) $\Rightarrow$ gives an asymptotic estimation of these bounds.
- Berry-Essen inequality $\Rightarrow$ gives a concrete precision for CLT.