

# Two Type-Theoretic Approaches to Probabilistic Termination

Ugo dal Lago   Charles Grellois

FOCUS Team – INRIA & University of Bologna

DICE-FOPARA 2017

# Motivations

- **Probabilistic** programming languages are more and more pervasive in computer science: modeling uncertainty, robotics, cryptography, machine learning, AI. . .
- **Quantitative** notion of termination: **almost-sure termination** (AST)
- AST has been studied for imperative programs in the last years. . .
- . . . but what about the probabilistic **functional** languages?

We introduce a **monadic, affine sized type system** sound for AST (our result at ESOP 2017), and sketch a **dependent, affine** type system for AST (work in progress).

## Sized Types: the Deterministic Case

Simply-typed  $\lambda$ -calculus is strongly normalizing (SN).

No longer true with the **letrec** construction. . .

**Sized types**: a **decidable** extension of the simple type system ensuring SN for  $\lambda$ -terms with letrec.

See notably:

- Hughes-Pareto-Sabry 1996, *Proving the correctness of reactive systems using sized types*,
- Barthe-Frade-Giménez-Pinto-Uustalu 2004, *Type-based termination of recursive definitions*.

# Sized Types: the Deterministic Case

Sizes:  $\mathfrak{s}, \mathfrak{t} ::= i \mid \infty \mid \widehat{\mathfrak{s}}$

+ size comparison underlying **subtyping**. Notably  $\widehat{\infty} \equiv \infty$ .

Idea:  $k$  successors = at most  $k$  constructors.

- $\text{Nat}^{\widehat{i}}$  is 0,
- $\text{Nat}^{\widehat{\widehat{i}}}$  is 0 or  $S\ 0$ ,
- ...
- $\text{Nat}^{\infty}$  is any natural number. Often denoted simply  $\text{Nat}$ .

The same for lists, ...

## Sized Types: the Deterministic Case

Sizes:  $\mathfrak{s}, \mathfrak{r} ::= i \mid \infty \mid \widehat{\mathfrak{s}}$

+ size comparison underlying **subtyping**. Notably  $\widehat{\infty} \equiv \infty$ .

Fixpoint rule:

$$\frac{\Gamma, f : \text{Nat}^i \rightarrow \sigma \vdash M : \text{Nat}^{\widehat{i}} \rightarrow \sigma[i/\widehat{i}] \quad i \text{ pos } \sigma}{\Gamma \vdash \text{letrec } f = M : \text{Nat}^{\mathfrak{s}} \rightarrow \sigma[i/\mathfrak{s}]}$$

“To define the action of  $f$  on size  $n + 1$ ,  
we only call recursively  $f$  on size at most  $n$ ”

# Sized Types: the Deterministic Case

Sizes:  $\mathfrak{s}, \mathfrak{t} ::= \mathfrak{i} \mid \infty \mid \widehat{\mathfrak{s}}$

+ size comparison underlying **subtyping**. Notably  $\widehat{\infty} \equiv \infty$ .

Fixpoint rule:

$$\frac{\Gamma, f : \text{Nat}^{\mathfrak{i}} \rightarrow \sigma \vdash M : \text{Nat}^{\widehat{\mathfrak{i}}} \rightarrow \sigma[\mathfrak{i}/\widehat{\mathfrak{i}}] \quad \mathfrak{i} \text{ pos } \sigma}{\Gamma \vdash \text{letrec } f = M : \text{Nat}^{\mathfrak{s}} \rightarrow \sigma[\mathfrak{i}/\mathfrak{s}]}$$

**Typable**  $\implies$  **SN**. Proof using reducibility candidates.

**Decidable** type inference.

# A Probabilistic $\lambda$ -calculus

$$M, N, \dots ::= V \mid V V \mid \text{let } x = M \text{ in } N \mid M \oplus_p N \\ \mid \text{case } V \text{ of } \{S \rightarrow W \mid 0 \rightarrow Z\}$$

$$V, W, Z, \dots ::= x \mid 0 \mid S V \mid \lambda x.M \mid \text{letrec } f = V$$

- Formulation equivalent to  $\lambda$ -calculus with  $\oplus_p$ , but constrained for technical reasons (A-normal form)
- Restriction to base type Nat for simplicity, but can be extended to general inductive datatypes (as in sized types)

# A Probabilistic $\lambda$ -calculus: Operational Semantics

$$\frac{}{\text{let } x = V \text{ in } M \rightarrow_v \{ (M[x/V])^1 \}}$$

$$\frac{}{(\lambda x.M) V \rightarrow_v \{ (M[x/V])^1 \}}$$

$$\frac{}{(\text{letrec } f = V) (c \vec{W}) \rightarrow_v \left\{ \left( V[f / (\text{letrec } f = V)] (c \vec{W}) \right)^1 \right\}}$$



# A Probabilistic $\lambda$ -calculus: Operational Semantics

$$\frac{}{\text{case } S \ V \text{ of } \{S \rightarrow W \mid 0 \rightarrow Z\} \rightarrow_v \left\{ (W \ V)^1 \right\}}$$

$$\frac{}{\text{case } 0 \text{ of } \{S \rightarrow W \mid 0 \rightarrow Z\} \rightarrow_v \left\{ (Z)^1 \right\}}$$

# A Probabilistic $\lambda$ -calculus: Operational Semantics

$$\frac{}{M \oplus_p N \rightarrow_v \{M^p, N^{1-p}\}}$$

$$\frac{M \rightarrow_v \{L_i^{p_i} \mid i \in I\}}{\text{let } x = M \text{ in } N \rightarrow_v \{(\text{let } x = L_i \text{ in } N)^{p_i} \mid i \in I\}}$$

# A Probabilistic $\lambda$ -calculus: Operational Semantics

$$\frac{\mathcal{D} \stackrel{VD}{=} \left\{ M_j^{p_j} \mid j \in J \right\} + \mathcal{D}_V \quad \forall j \in J, M_j \rightarrow_v \mathcal{E}_j}{\mathcal{D} \rightarrow_v \left( \sum_{j \in J} p_j \cdot \mathcal{E}_j \right) + \mathcal{D}_V}$$

For  $\mathcal{D}$  a distribution of terms:

$$\llbracket \mathcal{D} \rrbracket = \sup_{n \in \mathbb{N}} (\{ \mathcal{E}_n \mid \mathcal{D} \Rightarrow_v^n \mathcal{E}_n \})$$

where  $\Rightarrow_v^n$  is  $\rightarrow_v^n$  followed by projection on values.

We let  $\llbracket M \rrbracket = \llbracket \{ M^1 \} \rrbracket$ .

$M$  is AST iff  $\sum \llbracket M \rrbracket = 1$ .

# Random Walks as Probabilistic Terms

- **Biased** random walk:

$$M_{bias} = \left( \text{letrec } f = \lambda x. \text{case } x \text{ of } \left\{ S \rightarrow \lambda y. f(y) \oplus_{\frac{2}{3}} (f(SSy)) \mid 0 \rightarrow 0 \right\} \right) \eta$$

- **Unbiased** random walk:

$$M_{unb} = \left( \text{letrec } f = \lambda x. \text{case } x \text{ of } \left\{ S \rightarrow \lambda y. f(y) \oplus_{\frac{1}{2}} (f(SSy)) \mid 0 \rightarrow 0 \right\} \right) \eta$$

$$\sum \llbracket M_{bias} \rrbracket = \sum \llbracket M_{unb} \rrbracket = 1$$

Capture this in a sized type system?

## Another Term

We also want to capture terms as:

$$M_{nat} = \left( \text{letrec } f = \lambda x.x \oplus_{\frac{1}{2}} S (f x) \right) 0$$

of semantics

$$\llbracket M_{nat} \rrbracket = \left\{ (0)^{\frac{1}{2}}, (S 0)^{\frac{1}{4}}, (S S 0)^{\frac{1}{8}}, \dots \right\}$$

summing to 1.

(This is the **geometric distribution**.)

# Beyond SN Terms, Towards Distribution Types

**First idea:** extend the sized type system with:

$$\text{Choice} \quad \frac{\Gamma \vdash M : \sigma \quad \Gamma \vdash N : \sigma}{\Gamma \vdash M \oplus_p N : \sigma}$$

and “unify” types of  $M$  and  $N$  by **subtyping**.

Kind of **product interpretation** of  $\oplus$ : we can't capture more than SN...

# Beyond SN Terms, Towards Distribution Types

**First idea:** extend the sized type system with:

$$\text{Choice} \quad \frac{\Gamma \vdash M : \sigma \quad \Gamma \vdash N : \sigma}{\Gamma \vdash M \oplus_p N : \sigma}$$

and “unify” types of  $M$  and  $N$  by **subtyping**.

We get at best

$$f : \text{Nat}^{\hat{i}} \rightarrow \text{Nat}^{\infty} \vdash \lambda y. f(y) \oplus_{\frac{1}{2}} (f(SSy)) : \text{Nat}^{\hat{i}} \rightarrow \text{Nat}^{\infty}$$

and can't use a variation of the letrec rule on that.

# Beyond SN Terms, Towards Distribution Types

We will use **distribution types**, built as follows:

$$\text{Choice} \quad \frac{\Gamma | \Theta \vdash M : \mu \quad \Gamma | \Psi \vdash N : \nu \quad \{\mu\} = \{\nu\}}{\Gamma | \Theta \oplus_p \Psi \vdash M \oplus_p N : \mu \oplus_p \nu}$$

Now

$$f : \left\{ (\text{Nat}^i \rightarrow \text{Nat}^\infty)^{\frac{1}{2}}, \left( \text{Nat}^{\hat{i}} \rightarrow \text{Nat}^\infty \right)^{\frac{1}{2}} \right\}$$
$$\vdash$$
$$\lambda y. f(y) \oplus_{\frac{1}{2}} (f(SSy)) : \text{Nat}^{\hat{i}} \rightarrow \text{Nat}^\infty$$



## Designing the Fixpoint Rule

$$f : \left\{ (\text{Nat}^i \rightarrow \text{Nat}^\infty)^{\frac{1}{2}}, \left( \widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty \right)^{\frac{1}{2}} \right\}$$
$$\vdash$$
$$\lambda y. f(y) \oplus_{\frac{1}{2}} (f(SS y)) : \widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty$$

induces a random walk on  $\mathbb{N}$ :

- on  $n + 1$ , move to  $n$  with probability  $\frac{1}{2}$ , on  $n + 2$  with probability  $\frac{1}{2}$ ,
- on 0, loop.

The type system ensures that there is no recursive call from size 0.

Random walk AST (= reaches 0 with proba 1)  $\Rightarrow$  termination.

# Designing the Fixpoint Rule

$$\{\Gamma\} = \text{Nat}$$

$i \notin \Gamma$  and  $i$  positive in  $\nu$

$\{ (\text{Nat}^{s_j} \rightarrow \nu[i/s_j])^{p_j} \mid j \in J \}$  induces an AST sized walk

$$\text{LetRec} \frac{\Gamma \mid f : \{ (\text{Nat}^{s_j} \rightarrow \nu[i/s_j])^{p_j} \mid j \in J \} \vdash V : \text{Nat}^{\hat{i}} \rightarrow \nu[i/\hat{i}]}{\Gamma \mid \emptyset \vdash \text{letrec } f = V : \text{Nat}^{\tau} \rightarrow \nu[i/\tau]}$$

Sized walk: AST is checked by an external PTIME procedure.

# Generalized Random Walks and the Necessity of Affinity

A crucial feature: our type system is **affine**.

Higher-order symbols occur at most **once**. Consider:

$$M_{naff} = \text{letrec } f = \lambda x. \text{case } x \text{ of } \left\{ S \rightarrow \lambda y. f(y) \oplus_{\frac{2}{3}} (f(SSy)); f(SSy) \mid 0 \rightarrow 0 \right\}$$

The induced sized walk is AST, but  $M_{naff}$  is not.

# Key Property I: Subject Reduction

Main idea: reduction of

$$\emptyset \mid \emptyset \vdash 0 \oplus 0 : \left\{ \left( \text{Nat}^{\widehat{s}} \right)^{\frac{1}{2}}, \left( \text{Nat}^{\widehat{t}} \right)^{\frac{1}{2}} \right\}$$

is to

$$\left\{ \left( 0 : \text{Nat}^{\widehat{s}} \right)^{\frac{1}{2}}, \left( 0 : \text{Nat}^{\widehat{t}} \right)^{\frac{1}{2}} \right\}$$

- 1 Same **expectation type**:  $\frac{1}{2} \cdot \text{Nat}^{\widehat{s}} + \frac{1}{2} \cdot \text{Nat}^{\widehat{t}}$
- 2 Splitting of  $\llbracket 0 \oplus 0 \rrbracket$  in a typed representation  $\rightarrow$  notion of **pseudo-representation**

# Key Property I: Subject Reduction

## Theorem

Let  $M \in \Lambda_{\oplus}$  be such that  $\emptyset \mid \emptyset \vdash M : \mu$ . Then there exists a closed typed distribution  $\left\{ (W_j : \sigma_j)^{p'_j} \mid j \in J \right\}$  such that

- $\mathbb{E} \left( (W_j : \sigma_j)^{p'_j} \right) \preceq \mu$ ,
- and that  $\left[ (W_j)^{p'_j} \mid j \in J \right]$  is a pseudo-representation of  $\llbracket M \rrbracket$ .

By the soundness theorem of next slide, this inequality is in fact an equality.

## Key Property II: Typing Soundness

### Theorem (Typing soundness)

*If  $\Gamma \mid \Theta \vdash M : \mu$ , then  $M$  is AST.*

Proof by **reducibility**, using set of candidates parametrized by probabilities.

# Reducibility, the Probabilistic Case

Usual reducibility proof:

$M$  closed of type  $\sigma \Rightarrow M \in Red_\sigma \Rightarrow M$  is SN

In our setting:

# Reducibility, the Probabilistic Case

Usual reducibility proof:

$M$  closed of type  $\sigma \Rightarrow M \in \text{Red}_\sigma \Rightarrow M$  is SN

In our setting:

$$M \in \text{TRed}_\sigma^p \Rightarrow \sum \llbracket M \rrbracket \geq p$$



# Reducibility, the Probabilistic Case

Usual reducibility proof:

$M$  closed of type  $\sigma \Rightarrow M \in \text{Red}_\sigma \Rightarrow M$  is SN

In our setting:

$M$  closed of type  $\sigma \Rightarrow \forall p < 1, M \in \text{TRed}_\sigma^p \Rightarrow \forall p < 1, \sum \llbracket M \rrbracket \geq p$

$p$  increases with the number of fixpoint unfoldings we do, and we prove that  $M$  is in  $\text{TRed}_\sigma^p$  iff its  $n$ -unfolding is.

# Reducibility, the Probabilistic Case

Usual reducibility proof:

$M$  closed of type  $\sigma \Rightarrow M \in Red_\sigma \Rightarrow M$  is SN

In our setting:

$M$  closed of type  $\sigma \Rightarrow M \in TRed_\sigma^1 \Rightarrow \sum \llbracket M \rrbracket = 1$  i.e.  $M$  AST

by a **continuity** lemma.

## Reducibility, the Probabilistic Case – Open Terms

Usual case:  $\vec{x} : \vec{\sigma} \vdash M : \tau \Rightarrow \forall \vec{V} \in \overrightarrow{VRed}_{\vec{\sigma}}, M[\vec{x}/\vec{V}] \in Red_{\tau}$

## Reducibility, the Probabilistic Case – Open Terms

**Usual case:**  $\vec{x} : \vec{\sigma} \vdash M : \tau \Rightarrow \forall \vec{V} \in \overrightarrow{VRed_{\sigma}}, M[\vec{x}/\vec{V}] \in Red_{\tau}$

**In our setting:** if  $\Gamma \mid y : \{\tau_j^{p_j}\}_{j \in J} \vdash M : \mu$  then

- $\forall (q_i)_{i \in [0, 1]^n}, \forall \vec{V} \in \prod_{i=1}^n VRed_{\sigma_i}^{q_i},$
- $\forall (q'_j)_j \in [0, 1]^J, \forall W \in \bigcap_{j \in J} VRed_{\tau_j}^{q'_j},$
- we have  $M[\vec{x}, y/\vec{V}, W] \in TRed_{\mu}^{\alpha}$

where  $\alpha = \left( \prod_{i=1}^n q_i \right) \left( \left( \sum_{j \in J} p_j q'_j \right) + 1 - \left( \sum_{j \in J} p_j \right) \right).$

## Another Approach Using Dependent Types

Alternative approach to sized types: **dependent types**.

See Xi (2002), *Dependent Types for Program Termination Verification*.

Examples of dependent types à la Xi:

- $\varphi \mid \Gamma \vdash \underline{2} : int(2)$
- $\varphi \mid \Gamma \vdash \langle 2 \mid \underline{2} \rangle : \Sigma a : int. int(a)$

Terms of base type: annotated with size information which can be **packed** in the term (annotation by a size expression). Produces a **sum type** (existential).

$\varphi$ : context of constraints on free size variables, like  $a \in \{a \in int \mid a > 2\}$ .

## Another Approach Using Dependent Types

Alternative approach to sized types: **dependent types**.

See Xi (2002), *Dependent Types for Program Termination Verification*.

Examples of dependent types à la Xi:

- $\varphi \mid \Gamma \vdash + : \Pi \{a : \text{int}, b : \text{int}\} . \text{int}(a) \times \text{int}(b) \rightarrow \text{int}(a + b)$
- $\varphi \mid \Gamma \vdash \times : \Pi \{a : \text{int}, b : \text{int}\} . \text{int}(a) \times \text{int}(b) \rightarrow \text{int}(a \times b)$

Functions typically have universally quantified arguments (**product type**). Note that we could derive terms from  $+$  and  $\times$  which use sum types for return types.

## Another Approach Using Dependent Types

Sum types allow to get a uniform Choice rule:

$$\frac{\varphi | \Gamma | \Theta \vdash M : \sigma \quad \varphi | \Gamma | \Theta \vdash N : \sigma}{\varphi | \Gamma | \Theta \vdash M \oplus_p N : \sigma}$$

No longer need for distribution types!

Various sizes are annotated in the term.

## Another Approach Using Dependent Types

$$\frac{\begin{array}{l} \{\Gamma\} \subseteq \{\mathbf{bool}, \mathbf{int}\} \\ \varphi, \vec{a} : \vec{\tau} \mid \Gamma \mid f : \prod \vec{a} : \vec{\tau}. \sigma \Vdash_{\mathcal{P}} V : \sigma \\ \text{letrec}(\mathcal{P}, \rho) \text{ is AST for every } \rho \models \varphi \end{array}}{\varphi \mid \Gamma \mid \Theta \vdash \text{letrec } f[\vec{a} : \vec{\tau}] : \sigma = V : \prod \vec{a} : \vec{\tau}. \sigma}$$

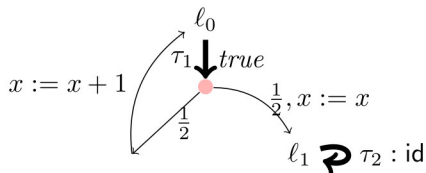
- Rely on PTS analysis (more general than random walks)
- letrec enters a **new mode**: typing relation  $\Vdash_{\mathcal{P}}$  indexed by a PTS

PTS = probabilistic transition system



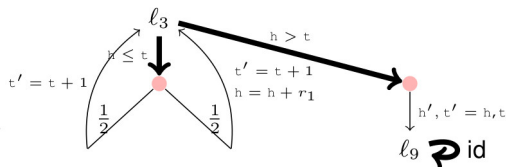
# Examples of PTS

```
1 int x := 0;  
2 while (flip (0.5))  
3   x ++;  
4 // end
```



Examples from Chakarav and Sankaranarayanan (2013), *Probabilistic Program Analysis with Martingales*

# Examples of PTS



Examples from Chakarov and Sankaranarayanan (2013), *Probabilistic Program Analysis with Martingales*

Our point: replaced sized walks by these processes modeling the flow of recursive calls. The process is built **on-the-fly** by the type system.

# Building the PTS

$$\frac{\varphi \vdash \vec{T} : \vec{\gamma} \quad \mathcal{P} = \mathbf{leaf} \left( f \left[ \vec{a} \mapsto \vec{T} \right] \right)}{\varphi \mid \Gamma \mid f : \prod \vec{a} : \vec{\gamma}. \sigma \Vdash_{\mathcal{P}} f \left[ \vec{T} \right] : \sigma[\vec{a}/\vec{T}]}$$

$\mathbf{leaf} \left( f \left[ \vec{a} \mapsto \vec{T} \right] \right)$  is a PTS with just one node, looping on itself and updating  $\vec{a}$  with  $\overrightarrow{[\![I]\!]_{\rho}}$ .

# Building the PTS

$$\frac{\begin{array}{l} \varphi \mid \Gamma \mid \emptyset \vdash M : \mathbf{bool}(I) \\ \varphi, I = 1 \mid \Delta \mid \Theta \Vdash_{\mathcal{P}} N : \sigma \\ \varphi, I = 0 \mid \Delta \mid \Theta \Vdash_{\mathcal{Q}} L : \sigma \end{array}}{\varphi \mid \Gamma, \Delta \mid \Theta \Vdash_{\text{if}(I, \mathcal{P}, \mathcal{Q})} \text{if } M \text{ then } N \text{ else } L : \sigma}$$

$\text{if}(I, \mathcal{P}, \mathcal{Q})$  is a PTS containing  $\mathcal{P}$  and  $\mathcal{Q}$  and with one new node branching to the root of  $\mathcal{P}$  or of  $\mathcal{Q}$  depending on  $\llbracket I \rrbracket_{\rho}$ .

# Building the PTS

$$\frac{\varphi | \Gamma | \Theta \Vdash_{\mathcal{P}} M : \sigma \quad \varphi | \Gamma | \Theta \Vdash_{\mathcal{Q}} N : \sigma}{\varphi | \Gamma | \Theta \Vdash_{\mathcal{P} \oplus_p \mathcal{Q}} M \oplus_q N : \sigma}$$

$\mathcal{P} \oplus_p \mathcal{Q}$  is a PTS containing  $\mathcal{P}$  and  $\mathcal{Q}$  and with one new node branching to the root of  $\mathcal{P}$  or of  $\mathcal{Q}$  depending on a biased coin flip of probability  $q$ .

# Building the PTS

$$\frac{\begin{array}{c} \{\Gamma\} \subseteq \{\mathbf{bool}, \mathbf{int}\} \\ \varphi, \vec{a} : \vec{\gamma} \mid \Gamma \mid f : \prod \vec{a} : \vec{\gamma}. \sigma \Vdash_{\mathcal{P}} V : \sigma \\ \text{letrec } (\mathcal{P}, \rho) \text{ is AST for every } \rho \models \varphi \end{array}}{\varphi \mid \Gamma \mid \Theta \vdash \text{letrec } f[\vec{a} : \vec{\gamma}] : \sigma = V : \prod \vec{a} : \vec{\gamma}. \sigma}$$

$\text{letrec } (\mathcal{P}, \rho)$  is a PTS obtained from  $\mathcal{P}$  by making the loops on the leaves pointing to the root of  $\mathcal{P}$ .

# Conjecture

We have strong hints that:

$$M \text{ has type } \sigma \Rightarrow M \text{ is AST.}$$

(we have a proof sketch based on the previous realizability argument).

Note that the system is again **affine**.

# Conclusion

First type system:

- **Affine** type system with **distributions** of types
- **Sized walks** induced by the letrec rule and solved by an external PTIME procedure
- **Subject reduction** + **soundness for AST**

Second type system:

- **Finer analysis**: more expressive sizes, modelization by PTS
- No need for distribution types thanks to sum types
- Still **affine**
- Soundness is work in progress

Thank you for your attention!



# Conclusion

First type system:

- **Affine** type system with **distributions** of types
- **Sized walks** induced by the letrec rule and solved by an external PTIME procedure
- **Subject reduction** + **soundness for AST**

Second type system:

- **Finer analysis**: more expressive sizes, modelization by PTS
- No need for distribution types thanks to sum types
- Still **affine**
- Soundness is work in progress

Thank you for your attention!

