

Distribution Types: A Type-Theoretic Approach To Almost Sure Termination

Ugo dal Lago Charles Grellois

FOCUS Team – INRIA & University of Bologna

Institute for Mathematical Sciences, Singapore
September 9, 2016

Motivations

- **Probabilistic** programming languages are important in computer science: modeling uncertainty, 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 **functional** probabilistic languages?

We introduce a **monadic, affine sized type system** sound for AST.

Sized types: the deterministic case

Simply-typed λ -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 λ -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 inducing **subtyping**. Notably $\widehat{\infty} = \infty$.

Idea:

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

The same for lists, ...

Sized types: the deterministic case

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

+ size comparison inducing **subtyping**. Notably $\widehat{\infty} = \infty$.

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.

Sized types: example in the deterministic case

From Barthe et al. (op. cit.):

$$\begin{aligned} \text{plus} \equiv & (\text{letrec } \text{plus}_{:\text{Nat}^s \rightarrow \text{Nat} \rightarrow \text{Nat}} = \\ & \lambda x_{:\text{Nat}^s}. \lambda y_{:\text{Nat}}. \text{case } x \text{ of } \{ \text{o} \Rightarrow y \\ & \quad \mid \text{s} \Rightarrow \lambda x'_{:\text{Nat}^s}. \text{s } \underbrace{(\text{plus } x' y)}_{:\text{Nat}} \\ & \quad \} \\ &) : \quad \text{Nat}^s \rightarrow \text{Nat} \rightarrow \text{Nat} \end{aligned}$$

Size decreases during recursive calls \Rightarrow SN.

A probabilistic λ -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 λ -calculus with \oplus_p , but constrained for technical reasons
- Restriction to base type `Nat` for simplicity, but can be extended to general inductive datatypes (as in sized types)

A probabilistic λ -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 λ -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 λ -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 λ -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:

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

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.

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 can't type M_{bias} nor M_{unb} in a way decreasing the size (essential for letrec): we get at best

$$f : \widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty \vdash \lambda y. f(y) \oplus_{\frac{2}{3}} (f(SSy)) : \widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty$$

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

$$f : \left\{ (\text{Nat}^i \rightarrow \text{Nat}^\infty)^{\frac{2}{3}}, \left(\widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty \right)^{\frac{1}{3}} \right\}$$
$$\vdash$$
$$\lambda y. f(y) \oplus_{\frac{2}{3}} (f(SSy)) : \widehat{\text{Nat}}^i \rightarrow \text{Nat}^\infty$$

Beyond SN terms, towards distribution types

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

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

Similarly:

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

Designing the fixpoint rule

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

$i \notin \Gamma$ and i positive in ν

??

LetRec

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

Designing the fixpoint rule

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

$i \notin \Gamma$ and i positive in ν

$$\sum_{j \in J} p_j(s_j) < 1$$

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

would allow to type

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

Designing the fixpoint rule

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

$i \notin \Gamma$ and i positive in ν

$$\sum_{j \in J} p_j(s_j) < 1 \text{ or } \sum_{j \in J} p_j < 1$$

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

would allow to type M_{nat} too:

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

Designing the fixpoint rule

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

$i \notin \Gamma$ and i positive in ν

$$\sum_{j \in J} p_j(\mathfrak{s}_j) < 1 \text{ or } \sum_{j \in J} p_j < 1$$

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

But how to cope with

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

Designing the fixpoint rule

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

$i \notin \Gamma$ and i positive in ν

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

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

solves the problem for

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

by deferring to an external **PTIME** procedure the convergence checking.

Generalized random walks and the necessity of affinity

A crucial feature: our type system is **affine**. Higher-order symbols occur at most **once**. Why? Consider:

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

and recall that its affine version was AST. Some reductions:

$$M_{naff} (S0) \rightarrow_v^* 0$$

$$M_{naff} (S0) \rightarrow_v^* M_{naff} (SS0); M_{naff} (SS0)$$

$$\rightarrow_v^* M_{naff} (S0); M_{naff} (SS0)$$

$$\rightarrow_v^* M_{naff} 0; M_{naff} (SS0)$$

$$\rightarrow_v^* 0; M_{naff} (SS0)$$

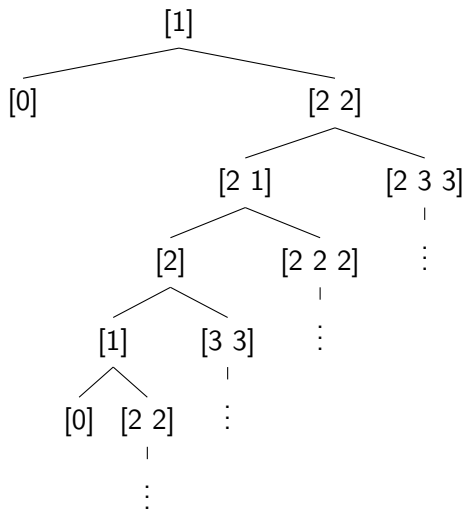
$$\rightarrow_v^* M_{naff} (SS0)$$

$$M_{naff} (S0) \rightarrow_v^* M_{naff} (SSS0); M_{naff} (SSS0); M_{naff} (SS0)$$

$$\rightarrow_v^* 0$$

Generalized random walks and the necessity of affinity

Tree of recursive calls:



Leftmost edges have probability $\frac{2}{3}$; rightmost ones $\frac{1}{3}$.

Generalized random walks and the necessity of affinity

Local shape

$$\begin{array}{c} [i_1 \cdots i_k] \\ \swarrow \quad \searrow \\ [i_1 \cdots i_k - 1] \quad [i_1 \cdots i_k + 1 \quad i_k + 1] \end{array}$$

when $i_k > 1$, and

$$\begin{array}{c} [i_1 \cdots i_{k-1} \quad 1] \\ \swarrow \quad \searrow \\ [i_1 \cdots i_{k-1}] \quad [i_1 \cdots i_{k-1} \quad 2 \quad 2] \end{array}$$

else. Leaves are all labeled with $[0]$.

The rightmost branch always increases the sum $i_1 + \cdots + i_k$ by at least 3
→ non AST random walk.

Key points

- **Affine** type system
- **Distribution** types
- **Sized walks** induced by the letrec rule and solved by an external PTIME procedure

Key properties

Theorem (Subject reduction)

Let $n \in \mathbb{N}$, and $\{(M_i : \mu_i)^{p_i} \mid i \in I\}$ be a closed typed distribution. Suppose that $\{(M_i)^{p_i} \mid i \in I\} \rightarrow_v^n \{(N_j)^{p'_j} \mid j \in J\}$ then there exists a closed typed distribution $\{(L_k : \nu_k)^{p''_k} \mid k \in K\}$ such that

- $\mathbb{E}((M_i : \mu_i)^{p_i}) = \mathbb{E}((L_k : \nu_k)^{p''_k})$,
- and that $[(L_k)^{p''_k} \mid k \in K]$ is a pseudo-representation of $\{(N_j)^{p'_j} \mid j \in J\}$.

Key properties

Theorem (Subject reduction)

Let $n \in \mathbb{N}$, and $\{(M_i : \mu_i)^{p_i} \mid i \in I\}$ be a closed typed distribution. Suppose that $\{(M_i)^{p_i} \mid i \in I\} \rightarrow_V^n \{(N_j)^{p'_j} \mid j \in J\}$ then there exists a closed typed distribution $\{(L_k : \nu_k)^{p''_k} \mid k \in K\}$ such that

- $\mathbb{E}((M_i : \mu_i)^{p_i}) = \mathbb{E}((L_k : \nu_k)^{p''_k})$,
- and that $[(L_k)^{p''_k} \mid k \in K]$ is a pseudo-representation of $\{(N_j)^{p'_j} \mid j \in J\}$.

Key properties

Theorem (Subject reduction)

Let $n \in \mathbb{N}$, and $\{(M_i : \mu_i)^{p_i} \mid i \in I\}$ be a closed typed distribution. Suppose that $\{(M_i)^{p_i} \mid i \in I\} \rightarrow_v^n \{(N_j)^{p'_j} \mid j \in J\}$ then there exists a closed typed distribution $\{(L_k : \nu_k)^{p''_k} \mid k \in K\}$ such that

- $\mathbb{E}((M_i : \mu_i)^{p_i}) = \mathbb{E}((L_k : \nu_k)^{p''_k})$,
- and that $[(L_k)^{p''_k} \mid k \in K]$ is a pseudo-representation of $\{(N_j)^{p'_j} \mid j \in J\}$.

Key properties

Theorem (Subject reduction)

Let $n \in \mathbb{N}$, and $\{(M_i : \mu_i)^{p_i} \mid i \in I\}$ be a closed typed distribution. Suppose that $\{(M_i)^{p_i} \mid i \in I\} \rightarrow_v^n \{(N_j)^{p'_j} \mid j \in J\}$ then there exists a closed typed distribution $\{(L_k : \nu_k)^{p''_k} \mid k \in K\}$ such that

- $\mathbb{E}((M_i : \mu_i)^{p_i}) = \mathbb{E}((L_k : \nu_k)^{p''_k}),$
- and that $\left[(L_k)^{p''_k} \mid k \in K \right]$ is a pseudo-representation of $\{(N_j)^{p'_j} \mid j \in J\}.$

Key properties

Theorem (Subject reduction)

Let $n \in \mathbb{N}$, and $\{ (M_i : \mu_i)^{p_i} \mid i \in I \}$ be a closed typed distribution. Suppose that $\{ (M_i)^{p_i} \mid i \in I \} \rightarrow_V^n \{ (N_j)^{p'_j} \mid j \in J \}$ then there exists a closed typed distribution $\{ (L_k : \nu_k)^{p''_k} \mid k \in K \}$ such that

- $\mathbb{E}((M_i : \mu_i)^{p_i}) = \mathbb{E}((L_k : \nu_k)^{p''_k}),$
- and that $\left[(L_k)^{p''_k} \mid k \in K \right]$ is a pseudo-representation of $\{ (N_j)^{p'_j} \mid j \in J \}.$

Key properties

Theorem (Typing soundness)

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

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

Next step: look for the type inference (decidable again??)

Thank you for your attention!

Key properties

Theorem (Typing soundness)

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

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

Next step: look for the type inference (decidable again??)

Thank you for your attention!

Key properties

Theorem (Typing soundness)

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

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

Next step: look for the type inference (decidable again??)

Thank you for your attention!