# Scott models for probabilistic computation:

## a Cartesian closed category for random variables

Pietro Di Gianantonio    Abbas Edalat

BCS Seminar, 26 March 2024

# Outline

- Develop a model of probabilistic computation based on Scott domains satisfying the following conditions:

(i) It is based on the fundamental notions of probability theory and denotational semantics

(ii) It supports a computable framework: an effective structure

(iii) It would give rise to the construction of simple and practical monads, in particular for functional programming languages, like Maybe, List, Powerset, etc.

# Modern Probability Theory

- Introduced by Andrey Kolomogoroff in 1933.
- It predates modern computer science.
- Fundamental basis of many branches of science
- Key role in Machine Learning, Deep Learning, Robotics, Quantum Computing, Modelling, Cognitive Science, etc.

# Sample Space and Probability space

- A **probability space** $(A, \Sigma_A, \nu)$ is a **sample space** $A$ with a $\sigma$-algebra $\Sigma$ of subsets or **events** (i.e., closed under countable union, intersection and complementation) and a **probability map** $\nu : \Sigma_A \to [0, 1]$, with $\nu(A) = 1$, and $\nu(\bigcup_{i \in \mathbb{N}} S_i) = \sum_{i \in \mathbb{N}} \nu(S_i)$ for disjoint events $S_i$'s.
- Usually $A = [0, 1]$ with $\Sigma_A$ generated by open intervals and the uniform (Lebesgue distribution) $\nu$, or,
- $A = \{0, 1\}^{\mathbb{N}}$ with $\Sigma_A$ generated by open cylinders sets $[x_1 x_2 \cdots x_n]$ with $x_i \in \{0, 1\}$ and $\nu([x_1 x_2 \cdots x_n]) = 2^{-n}$.

# Random Variables on Measurable Spaces

- A **measurable space** $X$ is a space with a $\sigma$-algebra $\Sigma_X$.
- A **random variable** on $X$ is a **measurable function** $r : A \to X$, i.e., with $r^{-1}(B) \in \Sigma_A$ for any $B \in \Sigma_X$.
- The **probability** of $B \in \Sigma_X$ is $\nu(r^{-1}(B))$, i.e., it is the probability of the event $r^{-1}(B)$ as determined by $\nu$.
- Two random variables $r_1, r_2 : A \to X$ are **equivalent** or have the same **probability distribution** if $\nu(r_1^{-1}(B)) = \nu(r_2^{-1}(B))$ for all $B \in \Sigma_X$.

# Domain Theory for Denotational Semantics

- Scott domains were introduced by Dana Scott in 1970:
- $D_\infty$ model of untyped lambda calculus $D_\infty \cong (D_\infty \to D_\infty)$
- Cartesian Closed Category: Hom$(X \times Y, Z) \cong$ Hom$(X, Z^Y)$
- Gordon Plotkin 1977: a simply typed lambda calculus:
- Programming Language for Computable Functions (PCF)

# Bounded complete dcpo with a finitary approximation

- In dcpo $(D, \sqsubseteq)$, the **way-below** or **finitary approximation**

  $b \ll a$ if $(c_i)_{i \in I}$ directed & $a \sqsubseteq \sup_{i \in I} c_i \Longrightarrow \exists i \in I. b \sqsubseteq c_i$.
- $B \subset D$ is a **basis** if

  $\forall a \in D. a = \sup_{i \in I}\{b \in B : b \ll a\}$ with the set directed.
- Countably based dcpo: **continuous** or **domain**, e.g., $\mathbb{N}_\perp$
- Basis of **Scott topology** denoted $\Omega D$:

  $\twoheaduparrow b = \{a \in D : b \ll a\}$ for $b \in B$.
- A **Scott domain** $D$ is a **bounded complete** domain, i.e, If

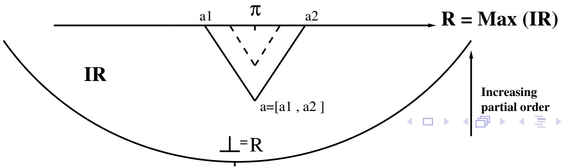
  any bounded set $S \subset D$ has its least upper bound $\bigsqcup S$.
- **Effective structure**: On basis elements $\ll$ is recursive.

# Scott domain **IR** of intervals of real numbers **R**

- **IR**: Bounded closed real intervals with reverse inclusion:
- Lub of a directed or bounded set of intervals:

  intersection of the intervals.
- Basis: rational intervals: $[b_1, b_2]$ with $b_1, b_2 \in \mathbb{Q}$.
- $[b_1, b_2] \ll [a_1, a_2]$ iff $b_1 < a_1$ and $a_2 < b_2$.
- Basis of Scott topology: For any rational interval $[b_1, b_2]$,

  $\twoheaduparrow[b_1, b_2] = \{[a_1, a_2] : [a_1, a_2] \subset \text{interior}([b_1, b_2])\}$.



# Function space: essential for CCC of Scott domains

- Suppose $D$ is a Scott domain and $X$ is a topological space with a continuous lattice of open sets, e.g., $X$ is a Scott domain or is any of the standard probability spaces $A$.
- Then the set of Scott continuous functions $(X \to D)$ ordered pointwise is itself a Scott domain.
- A **step function** $g : X \to D$ is of the form $g = \sup_{1 \leq i \leq n} d_i \chi_{O_i}$ with $O_i \subset X$ open and $d_i \in B_D$, and $g(x) = \sup\{d_i : x \in O_i\}$.
- Step functions provide a basis for $(X \to D)$.

# Monads for non-determinism

- Like the power set monad, there are three basic power domain monads for non-determinism for a domain $D$.
- **Lower (Hoare)** power domain: $LD$
- **Upper (Smyth)** power domain: $UD$
- **Convex (Plotkin)** power domain: $CD$
- In all these power domains, the basis consists of finite subsets of $D$ but with different ordering.
- The CCC of Scott domains is closed under $L$ and $U$.
- The CCC of “bifinite” domains is closed under C.

# Probabilistic power domain *PD* of a domain *D*

- Introduced by Saheb-Djahromi 79, Jones & Plotkin 89
- For space $X$, a **continuous valuation** $\sigma : \Omega X \to ([0,1], \leq)$ is a Scott continuous map on lattice of Scott opens of $X$: (i) $\sigma(\emptyset) = 0$, (ii) $\sigma(O_1 \cup O_2) + \sigma(O_1 \cap O_2) = \sigma(O_1) + \sigma(O_2)$.
- Simple valuations: $\sigma = \sum_{1 \leq i \leq n} q_i \delta_{b_i}$, where $b_i \in X$, $q_i \in \mathbf{R}$ with $\sigma(O) = \sum_{b_i \in O} q_i$, for open $O$.
- If $D$ is a domain, the space of continuous valuations $PD$ pointwise-ordered is a domain: simple valuations as basis.
- $P$ commutative monad on category of domains (Jones & Plotkin 89)
- Probability measures $\leftrightarrow$ continuous valuations (Alvarez et al. 2000)

# Open problem: Any CCC of domains closed under $P$?

- Classification of CCC of domains: Achim Jung 88, 89, 90
- No known CCC of domains closed under $P$ (Jung & Tix 98)
- Researchers had hoped such a CCC would provide a domain-theoretic model for probabilistic computation similar to power domains for non-determinism.

# Models lacking a CCC of domains closed under $P$

- Using other structures without an **effective struture.**
- CCC of dcpo's with commutative monad $P_0$, where $P_0 C$ is the smallest dcpo containing simple valuations on dcpo $C$.

  Jia, Lindenhovius, Mislove, Zamdzhiev, 21. And Goubault-Larrecq, Jia, Théron, 23

  Unlike a domain, $P_0 C$ has no effective structure. Also:

  Given $f : C \to P_0 E$, binding operation requires

  $f^\dagger : P_0 C \to P_0 E$ with $(f^\dagger(q))(O) = \int_C (f(c))(O)\ dq(c)$,

  generally non-computable.
- Other models with non-standard, new concepts, structures:

  Quasi-Borel spaces Vákár, Kammar, Staton 2019

  Probabilistic coherent spaces Danos, Ehrhard 2011

# A natural approach to probabilistic computation

- Scott domains provide the basic and elegant structure in modelling non-probabilistic computation successfully
- Opens as observable events: Smyth, Abramsky, Vickers, Jung
- Random variables—from standard sample/probability spaces—on Scott domains capture probability distributions
- This **function representation** of probability distributions is aligned with the **functional language paradigm.**
- Thus, a simple data type for probabilistic computation is provided by random variables on Scott domains with an equivalence relation on random variables

# Representation theorem for probability distributions

- $A = \{0,1\}^{\mathbb{N}}$ or $[0,1]$ with uniform probability distribution $\nu$.
- Let $D$ be a Scott domain. Then so is $(A \to D)$.
- $P^1 D$ normalised probabilistic power domain of $D$ (E. 1994)
- **Theorem** $T : (A \to D) \to P^1 D$ with $(T(r))(O) = \nu(r^{-1}(O))$
- $T$ is a continuous surjection (surjection: Mislove 2016)
- $T$ takes step functions to simple valuations:

  $T(\sup_{1 \le i \le n} d_i \chi_{O_i}) = \sum_{1 \le i \le n} \nu(O_i)\delta(d_i)$
- $T$ is an **effectively given map** both ways
- Preserves $\ll$ for $A = (0,1)$, $\{0,1\}^{\mathbb{N}}_*$ (no infinite recurring 0)

# Domains with a Partial Equivalence Relation PER

- A **PER domain** $(D, \sim)$ is a Scott domain $D$ with a symmetric and transitive **logical** relation $\sim$ satisfying: (i) $\bot \sim \bot$ and (ii) for chains $(d_i)_{i\in\mathbb{N}}$ and $(d'_i)_{i\in\mathbb{N}}$, $\forall i \in \mathbb{N}\; d_i \sim d'_i \Longrightarrow \sup_{i\in\mathbb{N}} d_i \sim \sup_{i\in\mathbb{N}} d'_i$
- Equivalence of random variables satisfies these conditions.
- For $f, g : D \to E$, define $f \sim g$ if $d \sim d' \Longrightarrow f(d) \sim g(d')$
- Morphism $[f] : (D, \sim) \to (E, \sim)$ as PER class of maps
- $((D, \sim_D) \to (E, \sim_E))$ is defined as $(D \to E)$ with $\sim_{D\to E}$ defined as PER of maps. Similarly, $(D, \sim_D) \times (E, \sim_E)$.
- **PER**: The category of Scott domains with PER is a CCC

# Random variable functor *R* on **PER** category

- The **R-topology** $\Upsilon_D \subset \Omega D$ on a PER domain $(D, \sim_D)$, as a sub-Scott topology, consists of Scott open sets $O \subset D$ closed under $\sim_D$, i.e., $(d \in O \,\&\, d' \sim_D d) \implies d' \in O$.
- The *R* **functor** is defined on **PER** category by:
- On objects: $R(D, \sim_D) = ((A \to D), \sim_{A \to D})$ with $r \sim_{A \to D} s$ if $\forall \omega \in A.\, r(\omega) \sim_D r(\omega) \,\&\, s(\omega) \sim_D s(\omega)$,and $\forall O \in \Upsilon_D.\, \nu(r^{-1}(O)) = \nu(s^{-1}(O))$.
- On morphisms: $[f] : (D, \sim_D) \to (E, \sim_E)$ with $R[f] : ((A \to D), \sim_{RD}) \to ((A \to E), \sim_{RE})$ given by $R[f] = [\lambda r.\, f \circ r]$.

# Turning random variable functor $R$ into a monad

- Let $h : (A, \nu) \to (A^2, \nu \times \nu)$ be a measure-preserving continuous bijection on a set of full $\nu$-measure on $A$.
- There are infinitely many ways to choose $h$. For example:
- For $\omega \in A = \{0, 1\}^{\mathbb{N}}$, define $h(\omega) = (\omega^{e}, \omega^{o})$, where $\omega^{e}$ (respectively, $\omega^{o}$) is the sequence of values in even (respectively, odd) positions in $\omega$, i.e., for $\in \mathbb{N}$
- $(\omega^{e})_i = \omega_{2i}$ and $(\omega^{o})_i = \omega_{2i+1}$
- $h$ is a homeomorphism with inverse $k := h^{-1} : A^2 \to A$: $k(\omega, \omega') = \omega''$, where $\omega''_{2i} = \omega_i$ and $\omega''_{2i+1} = \omega'_i$ for $i \in \mathbb{N}$.

# A monad using the Hilbert space filling curve

- For $A = [0,1]$, let $h : [0,1] \to [0,1]^2$ be the Hilbert curve.
- Domain-theoretic representation by four affine maps

  $H_i : [0,1]^2 \to [0,1]^2$, where $i = 0, 1, 2, 3$:

$$H_0 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad H_1 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$H_2 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

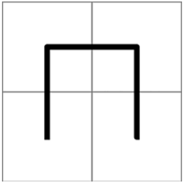
$$H_3 \begin{pmatrix} x \\ y \end{pmatrix} = \frac{-1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

- For $\omega \in [0,1]$, using quaternary representation, define:

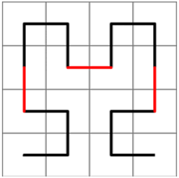
  $h(\omega) = h(0._4\omega_0\omega_1\omega_2\ldots) = \bigcap_{i\in\mathbb{N}} H_{\omega_0} H_{\omega_1} \ldots H_{\omega_i}([0,1]^2)$

# Domain-theoretic generation of Hilbert's curve, E. 93

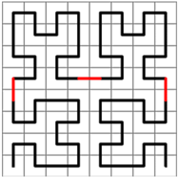
- Three iterates starting with $S = [0,1]^2$ of the map of sub-squares: $H : \mathbf{I}[0,1]^2 \to \mathbf{I}[0,1]^2$ with $H(S) = \bigcup_{i=0}^{3} H_i[S]$



$H([0,1]^2)$



$H^2([0,1]^2)$



$H^3([0,1]^2)$

- The $n^{\text{th}}$ iterate $H^n([0,1]^2)$ gives a grid of $4^n$ sub-squares.
- $h$ is 1-1 almost everywhere, with the countable exceptions: four-to-one on any grid node, two-to-one on any grid line
- Easy derivation of $h_1(\omega)$, $h_2(\omega)$ for $\omega = 0._4\omega_0\ldots$ with $h = \langle h_1, h_2\rangle$.

# Very simple monads for probabilistic computation

- Probability space $(A, \nu)$ for $A = \{0,1\}^{\mathbb{N}}$ or $A = [0,1]$.
- Monad $R$ : **PER** $\to$ **PER** with natural transformations
- **Unit:** $\eta_D : D \to (A \to D)$ with $\eta_D(d)(\omega) = d$.
- **Flattening:** $\mu_D : (A \to (A \to D)) \to (A \to D)$ with

  $\mu_D(r)(\omega) = r(h_1(\omega))(h_2(\omega))$.
- Alternatively, as a Kleisli triple we have $(R, \eta, (-)^\dagger)$ with

  $(-)^\dagger : (D \to (A \to E)) \to ((A \to D) \to (A \to E))$ given by:

  $f^\dagger(r)(\omega) = f(r(h_1(\omega)))(h_2(\omega))$.
- **Note:** Monadic properties hold up to equivalence.

# The monads are strong and commutative

- Useful monads are both strong and commutative.
- For $\langle D, \sim_D \rangle, \langle E, \sim_E \rangle \in$ **PER**, define

  $t_{D,E} : D \times RE \to R(D \times E)$, by $t_{D,E}(d, s) = \langle \eta_D(d), s \rangle$.
- This gives a tensorial strength, i.e., $R$ is strong.
- Also, $t'_{D,E} : RD \times E \to R(D \times E)$ with $t'_{D,E}(r, e) = \langle r, \eta_E(e) \rangle$,

  $t'^{\dagger}_{D,E} \circ t_{RD,E}(r, s) = t'^{\dagger}_{D,E} \langle \eta_D(r), s \rangle = \langle \eta_D(r), \eta_E(s) \rangle$,

  $t^{\dagger}_{D,E} \circ t'_{D,RE}(r, s) = t^{\dagger}_{D,E} \langle r, \eta_E(s) \rangle = \langle \eta_D(r), \eta_E(s) \rangle$.
- As the two expressions are equal, $R$ is commutative.

# IID random variables for any probability distribution

- The inverse transform $G$ of the cumulative distribution of any probability distribution $P$ on $\mathbf{R}$ gives a random variable $G^* : [0,1] \to \mathbf{IR}$ with support on $\mathbf{R} = \mathrm{Max}(\mathbf{IR})$.
- Any random variable $r : A \to D$ gives IID $h_1(r)$ and $h_2(r)$.
- The monad gives $2^n$ IID random variables equivalent to $r$: $h_{i_n}(h_{i_{n-1}}(\ldots(h_{i_2}(h_{i_1}(r)))\ldots))$ for $1 \leq i_t \leq 2$ with $1 \leq t \leq n$.
- EG. By Box-Muller transform, two **Gaussian** IID on $\mathbf{R}$:

  $z_1 = \sqrt{-2\ln h_1}\cos 2\pi h_2 \qquad z_2 = \sqrt{-2\ln h_2}\cos 2\pi h_1$
- For Gaussian IID $r_i$'s, **student t-distribution** of degree $n$:

  $r = \frac{r_{n+1}}{\sqrt{n^{-1}\sum_{i=1}^{n} r_i^2}} : [0,1] \to \mathbf{IR}$, supported on $\mathbf{R}$.

# Functions of random variables: Dirichlet distribution

- For $\alpha_i > 0$ and random variables $x_i$ on $[0,1]$ with $1 \leq i \leq k$,

  $B(\alpha) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)}$ Dirichlet distribution is: $\frac{1}{B(\alpha)} \prod_{1 \leq i \leq k} x_i^{\alpha_i - 1}$
- The domain-theoretic Dirichlet distribution

  $D_\alpha : (A \to \mathbf{I}[0,1])^k \to (A \to \mathbf{I}[0,1])$ given by

  $D_\alpha(r_1, \ldots, r_k) = \lambda\omega. \frac{\prod_{i=1}^{k} (r_i(\omega))^{\alpha_i - 1}}{B(\alpha)}$
- **Pointwise extension of the power map**:

  $x \mapsto x^a : [0,1] \to [0,1]$ is $x \mapsto x^a : \mathbf{I}[0,1] \to \mathbf{I}[0,1]$ with

$$x^a = \begin{cases} [(x^-)^a, (x^+)^a] & a \geq 0 \\ [(x^+)^a, (x^-)^a] & a < 0 \end{cases}$$

# Further work

- A Simply typed $\lambda$-calculus for higher order probabilistic computation with a constant for sampling
- Haskell implementation
- Conditional probabilities and Bayesian statistics
- Expectation values of functions of random variables
- Probabilistic computation on Polish (complete separable metrizable) spaces
- Domain-theoretic model of stochastic processes