On Random Variable Models of Domains

Michael W. Mislove

Tulane University
New Orleans, LA

Sixth International Symposium on
Domain Theory and Its Applications
Hunan University
October, 2013

Work Supported by NSF and US AFOSR

Some heartfelt thanks

To President Zhao, Deans Jiang and Li, Professor Lu, Professor Guo-Qiang Zhang

And to all those who worked behind the scenes to make this such a great success!

Thank You Very Much!!

Probability is fundamental for computational models

Two approaches:

- Randomized computation over a predefined, parameterized family of measurable sets
  - Dana's *Stochastic Lambda Calculus*
    - Randomized algorithms

Versus

- Finding mechanisms to model probability within arbitrary domains

Probability is fundamental for computational models

- Probability is fundamental security
  - Basis for definition
  - Participants in crypto-protocols make random choices
  - Quantitative information flow uses entropy, capacity and related statistics

Models used to analyze system security must support reasoning about probability.

But probabilistic domain models are difficult...

The Probabilistic Power Domain and Its Problems

- The Probabilistic Power Domain
  - $\mathsf{SProb}(D)$ – subprobability measures over D form a domain
    - $\mu \leq \nu$ iff $\mu(O) \leq \nu(O)$ ($\forall O$ open)
  - Extends order on underlying domain under $x \mapsto \delta_x \colon D \to \mathsf{SProb}(D)$
  - Forms monad on DCPO, on Dom and on CohDom
- Powerful model for reasoning about specification and refinement
  - Morgan, McIver, et al apply the probabilistic power domain to the traditional CSP models.
- Doesn't play well with other monads:
  - No distributive law wrt nondeterminism monads
- No known invariant Cartesian closed category of domains
- Shortcomings led to search for alternative models

Traditional model of random choice

- Basic model is *binary choice:* $p +_r q,\ \ r \in [0,1]$
  - Flips of a (fair?) coin...
- As computation evolves, choices generate *trace distributions*
  - Idea taken from trace models of process calculi
- Start with probabilistic automaton $S \xrightarrow{flip} \mathrm{Prob}(\{0,1\} \times S)$,
  - Begin in start state – δ_{s_0}, then evolve to
  - $r\delta_{(0,s_0 s_1)} + (1-r)\delta_{(1,s_0 s_2)}$

    $\vdots$
  - $\sum_{i=1}^{2^n} r_i \delta_{(b_0 \cdots b_{n-1},\alpha_i)}, \quad \sum_i r_i = 1, \alpha_i \in S$

    $\vdots$
  - Natural model is $\mathrm{Prob}((\{0,1\} \times S)^\infty)$.
- $\mathrm{Prob}((\{0,1\} \times S)^\infty)$ is bounded complete, but for more complicated domains D, $\mathrm{Prob}(D)$ poorly understood.

  Random variables offer an alternative

Random variables

- Let (X, Σ, μ) be a probability space with probability measure μ.

  A *random variable* on X is a measurable function $f : X \to Y$, where Y is a measure space.
- Take X and Y to be domains, f Scott continuous
- *Idea:* Choose X a "standard domain" satisfying $\mathit{Prob}(X)$ is a "nice" domain.

  Then: model of random variables on Y is $\mathit{Prob}(X) \times [X \to Y]$

  Stays in any Cartesian closed category containing Y.

An Example

In case of $S \xrightarrow{\mathit{flip}} \mathrm{Prob}(\{0,1\} \times S)$

- $D = S^\infty$
- μ_n is the measure on $\{0,1\}^n$ generated by flipping the coin $\delta_0 +_r \delta_1$ n times, and
- $f_n \colon \{0,1\}^n \to S^\infty$ by $f_n(b_0 \cdots b_{n-1}) = \alpha_n$, the chosen element depending on the outcome of the n possible flips.
- $\delta_\epsilon = \mu_0 \leq \mu_1 \leq \cdots \leq \mu_n \leq \mu_{n+1} \leq \cdots$

Simple Random Variable Model

- Use *Cantor Tree* $\mathcal{C} \simeq \{0,1\}^* \cup \{0,1\}^\omega$ for standard domain.
- Simple Random Variable domain $SRV(D)$:

  $\{(\mu_n, f_n) \mid \mu_n \in \mathrm{Prob}(2^n) \text{ and } f_n \colon 2^n \to D\}$

  $(\mu_m, f_m) \leq (\mu_n, f_n)$ iff $m \leq n, \pi_{2^m}(\mu_n) = \mu_m$ and $f_m \circ \pi_{2^m} \leq f_n$
- Model first proposed by *Goubault-Larrecq & Varacca*

Random Variable model (cont'd)

- $SRV(D) \equiv \bigoplus_n \mathrm{Prob}(2^n) \times [2^n \to D]$ is a monad:

$$\begin{tikzcd}
SRV(D) \arrow[drr, bend left, "h^\dagger"] & & \\
& D \arrow[ul, hook, "\eta"] \arrow[r, "h"'] & SRV(E)
\end{tikzcd}$$

  $\eta(x) = (\delta_\epsilon, \mathrm{const}_x)$

- *Problem:* $h^\dagger$ is not monotone!
- Originates from viewing successive coin flips as increasing in the order...

Alternative model I

- *Basic idea:* Flatten model so concatenation doesn't need to be monotone in first component.
- Leads to model which looks like

$$1 \text{ flip} \oplus 2 \text{ flips} \oplus 3 \text{ flips} \oplus \cdots \oplus n \text{ flips} \oplus \cdots$$

- Begin with $\mathrm{SProb}(n) = \left\{\sum_{i<n} r_i\delta_i \mid 0 \leq r_i \;\&\; \sum_i r_i \leq 1\right\}$
  - $\sum_i r_i\delta_i \leq \sum_i s_i\delta_i$ iff $r_i \leq s_i$ $(\forall i)$.
  - $\sum_i r_i\delta_i \wedge \sum_i s_i\delta_i = \sum_i (r_i \wedge s_i)\delta_i$
  - $\perp = 0$
  - A directed $\Rightarrow (\sup A)(i) = \sup_{\mu \in A} \mu(i)$.

Alternative model I

- Flat random variable domain:

  $$RV^\flat(D) = \bigoplus_n \left(\mathrm{SProb}(2^n) \times D^{2^n}\right)$$

  - $(\mu_n, X_n) \leq (\nu_m, X_m)$ iff $m = n, \mu_n \leq \nu_n$, and $X_n(i) \leq X_m(i)$ $(\forall i)$.
- $f\colon D \to E \;\Rightarrow\; RV^\flat(f)\colon RV^\flat(D) \to RV^\flat(E)$

  by $RV^\flat(f)(\mu_n, X_n) = (\mu_n, f \circ X_n)$.
- $RV^\flat(D)$ forms a monad on BCD.
- *Problem:* $RV^\flat(D)$ makes too many distinctions:

  $(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_1, (a, b)) \neq (\frac{2}{3}\delta_0 + \frac{1}{3}\delta_1, (b, a))$, etc.
- Solution requires some background work.

Free ordered semigroup

- $P^* = \bigoplus_{n>0} P^n$ is free ordered semigroup over poset P:
  - $w \leq w'$ iff $|w| = |w'|$ & $w_i \leq_P w'_i$ $(\forall i \leq |w|)$.
  - $ww' \in P^{m+n}$ if $w \in P^m$ & $w' \in P^n$.
  - *Note (J.-E. Pin):* Free ordered *monoid* is flat.
- Also works for P in BCD, FS, or RB.
- To obtain the free *commutative* semigroup, we take a quotient:
  - $S(n)$ acts on P^n by permuting the components.
  - $P^n/S(n)$ is the set of n-bags over P.
  - $\pi_n \colon P^n \to P^n/S(n)$ is monotone.
- $COS(P) = \bigoplus_{n>0} P^n/S(n)$ – free commutative ordered semigroup over P.

Free ordered domain

- Rudin's Lemma implies this also works in domains.
- $CDS(P) = \bigoplus_{n>0} P^n/S(n)$ – free commutative domain semigroup over domain P.

Apply this to $RV^\flat(D)$ to obtain flat "commutative" random variable domain:

- $CRV^\flat(D) = \bigoplus_n \left(\mathrm{SProb}(2^n) \times D^{2^n}\right)/S(2^n)$
- Now $(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_1, (a, b)) \equiv (\frac{2}{3}\delta_0 + \frac{1}{3}\delta_1, (b, a))$, etc.
- *But:* $(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_1, (a, b)) \not\equiv (\frac{1}{3}\delta_0 + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_1, (a, b, b))$
- Still a monad over RB and FS (but not over BCD).

Some Additional Comments

- Work was inspired by Varacca's *indexed valuations* (2004) and Goubault-Larrecq & Varacca's work on the first model.
- Jean Goubault-Larrecq is working on a patch to the first model.
  - Refines the order
- Tyler Barker also working on a patch
  - Redefines the Kleisli lift – somewhat akin to conditional probability
- Remaining question: Can the second model be extended to include recursion on the number of flips?