Our sponsors make OPLSS possible.

Diamond Sponsor

Jane Street

https://www.janestreet.com/

Platinum Sponsors

National Science Foundation
WHERE DISCOVERIES BEGIN

https://www.nsf.gov/

SIGPLAN

https://www.sigplan.org/

Silver Sponsors

OREGON PROGRAMMING LANGUAGES SUMMER SCHOOL

Introduction to Programming Language Semantics

Lecture 4: Axiomatic and Probabilistic Semantics

Roadmap

- Lecture 1: Structural operational semantics
  IMP language, small-step rules, big-step rules, evaluation contexts, big-step rules as binary relations, algebraic properties
- Lecture 2: The λ-calculus
  λ-terms, safe substitution, binding and scope, α- and β-reduction, Church-Rosser theorem, data encodings, recursion, simple types, progress and preservation
- Lecture 3: Denotational semantics
  CPOs, least fixpoints, Knaster-Tarski theorem, domain equations
- Lecture 4: Axiomatic semantics, probabilistic semantics
  Weakest preconditions, semantics of Hoare logic, Dynamic logic, probabilistic semantics

Styles of Semantics

Three Styles of Semantics

- Operational: Focus on states, local, forward-moving
- Denotational: Focus on states, global, forward-moving
- Axiomatic: Focus on observations, backward-moving

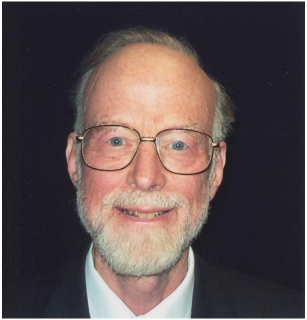
Axiomatic Semantics

What can you observe about states?

- Let S be a set of states
  e.g., for IMP, the states are environments Env $=$ Var $\to \mathbb{N}$
- Tests are typically some syntactically defined set of properties that are either true or false of a state — these determine what we can observe
  e.g., for IMP, we had BExp:
  $b \ ::= \ a_1 \leq a_2 \mid a_1 = a_2 \mid b_1 \wedge b_2 \mid \neg b \mid \ldots$
- The semantics determines a satisfaction relation between states and tests — write $s \vDash B$ when s satisfies the test B
  e.g., for IMP, $s \vDash a_1 \leq a_2$ if $\langle a_1 \leq a_2, s \rangle \!\downarrow_b \mathtt{true}$

Hoare logic

- Hoare logic $\{B\}\, p\, \{C\}$
  "If the precondition B is observed immediately before execution, and if p halts, then the postcondition C will be observed immediately after execution"
- Viewed as a means of specifying the meaning of programs



C. A. R. "Tony" Hoare
(1934–2026)

Rules of Hoare logic

$$\frac{}{\{B[a/x]\}\; x := a\; \{B\}}$$

$$\frac{}{\{B\}\; \texttt{skip}\; \{B\}}$$

$$\frac{\{B\}\, p\, \{C\} \quad \{C\}\, q\, \{D\}}{\{B\}\, p\,;q\, \{D\}}$$

$$\frac{\{B \wedge C\}\, p\, \{D\} \qquad \{\neg B \wedge C\}\, q\, \{D\}}{\{C\}\; \texttt{if}\; B\; \texttt{then}\; p\; \texttt{else}\; q\; \{D\}}$$

$$\frac{\{B \wedge C\}\, p\, \{C\}}{\{C\}\; \texttt{while}\; B\; \texttt{do}\; p\; \{\neg B \wedge C\}}$$

Relative completeness

Theorem (S. Cook, 1974)

Hoare logic is relatively complete.

- Relatively complete means able to prove all valid partial correctness assertions given an oracle for all properties of the domain of computation expressible in the underlying assertion language
- the assertion language must be able to express the weakest liberal precondition of any given program and postcondition

Weakest (liberal) preconditions

The weakest liberal precondition of a program p and a postcondition B (denoted $\mathsf{wlp}(p, B)$) is set of all states that guarantee that if p halts, then the halting state will satisfy B

$$
\begin{aligned}
&\mathsf{wlp}(\mathtt{skip}, B) = B \\
&\mathsf{wlp}(x := a, B) = B[a/x] \\
&\mathsf{wlp}(p\,;q, B) = \mathsf{wlp}(p, \mathsf{wlp}(q, B)) \\
&\mathsf{wlp}(\mathtt{if}\ B\ \mathtt{then}\ p\ \mathtt{else}\ q, C) \\
&\qquad = (B \Rightarrow \mathsf{wlp}(p, C)) \wedge (\neg B \Rightarrow \mathsf{wlp}(q, B)) \\
&\text{if } B \wedge I \Rightarrow \mathsf{wlp}(p, I) \text{ and } \neg B \wedge I \Rightarrow C \\
&\qquad \text{then } I \Rightarrow \mathsf{wlp}(\mathtt{while}\ B\ \mathtt{do}\ p, C)
\end{aligned}
$$

The weakest precondition of p and B (denoted $\mathsf{wp}(p, B)$) is set of all states that guarantee that p will halt, and the halting state will satisfying B — $\mathsf{wp}(p, B) \Leftrightarrow \mathsf{wlp}(p, B) \wedge$ termination

Dynamic Logic (V. Pratt 1976)

Modal logic with programs as modalities

- $s \models [p]A \;\Leftrightarrow\; \forall t\,(s,t) \in [\![p]\!] \Rightarrow t \models A$
- $s \models \langle p \rangle A \;\Leftrightarrow\; \exists t\,(s,t) \in [\![p]\!] \land t \models A$

Axioms

Axioms of propositional logic	$[p+q]A \Leftrightarrow [p]A \land [q]A$
$[p]A \Leftrightarrow \neg\langle p \rangle \neg A$	$[p;q]A \Leftrightarrow [p][q]A$
$[p](A \Rightarrow B) \Rightarrow [p]A \Rightarrow [p]B$	$[p^*]A \Leftrightarrow A \land [p][p^*]A$
$[A]B \Leftrightarrow A \Rightarrow B$	$(A \land [p^*](A \Rightarrow [p]A)) \Rightarrow [p^*]A$

Rules of inference:

$$\frac{A \qquad A \Rightarrow B}{B} \qquad\qquad \frac{A}{[p]A}$$

Axiomatic Semantics

Predicate transformers

$\langle\ \rangle$, $[\ ]$ of dynamic logic and wp/wlp are called predicate transformers because they map postconditions to preconditions

Backward in time!

Some relations:

- $A \Rightarrow [p]B \;\Leftrightarrow\; \{A\}\, p\, \{B\}$
- $[p]B = \mathsf{wlp}(p, B)$

Duality

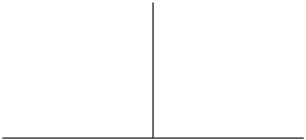
$$
\begin{tikzcd}
s \models [p]A \arrow[rr, bend left, "\mathcal{C}[\![p]\!]"] & \Leftrightarrow & \mathcal{C}[\![p]\!]s \models A \arrow[ll, bend left, "{[p]}"]
\end{tikzcd}
$$

Probabilistic Semantics

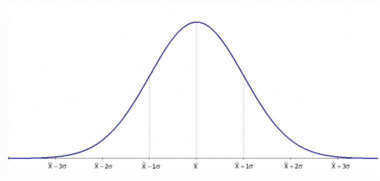
Deterministic reasoning tends to be more truth-functional

- Correctness: Is the postcondition always satisfied?
- Termination: Does this program halt on all inputs?
- Fair termination: Does this concurrent program always terminate, provided scheduling is fair?
- Worst-case complexity: Does this program always halt within $O(n^2)$ time?

Probabilistic Semantics



$\Rightarrow$



Probabilistic reasoning tends to be more quantitative

- Correctness: What is the probability that the postcondition is satisfied?
- Termination: What is the likelihood that this program halts?
- Average-case complexity: What is the expected halting time, given this input distribution on inputs of length n?

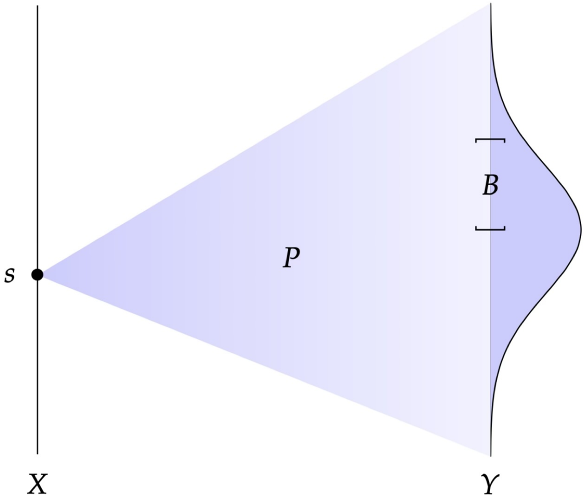
Adding Probability

$$\begin{array}{rcl}
\text{states} & \Rightarrow & \text{measures} \\
\text{predicates} & \Rightarrow & \text{measurable functions} \\
\models & \Rightarrow & \int \\
\text{binary relations} & \Rightarrow & \text{Markov kernels} \\
\text{predicate transformers} & \Rightarrow & \text{measurable function transformers}
\end{array}$$

Adding probability to IMP

- Add a random assignment $x := \texttt{rnd}$
- Each call to $x := \texttt{rnd}$ assigns a random value to x
- Samples are chosen according to some fixed probability distribution
- Samples are independent
- Examples:
  - $x := \texttt{rnd}$ assigns 0 or 1 to x, each with probability 1/2 (fair coin)
  - $x := \texttt{rnd}$ assigns a random real in $[0, 1]$ to x according to the uniform distribution on $[0, 1]$ — the sample will be in the interval $[a, b]$ with probability $b - a$

Markov Kernels



Measurable Spaces and Functions

- The state space must be a measurable space $(S, \mathcal{B})$, where S is a set and $\mathcal{B} \subseteq 2^S$, the measurable sets or events
- Elements of $\mathcal{B}$ are the observable events — they are the sets that can have a probability — all tests must denote a set in $\mathcal{B}$
- $\mathcal{B}$ must form a σ-algebra on S — a Boolean algebra closed under countable union
- A function $f : (S_1, \mathcal{B}_1) \to (S_2, \mathcal{B}_2)$ is measurable if the inverse image of any measurable set in $\mathcal{B}_2$ is measurable in $\mathcal{B}_1$

Examples

- Discrete space $(A, 2^A)$, A any finite or countable set
- $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ are the Borel sets, the smallest σ-algebra on $\mathbb{R}$ containing all intervals
- $([0,1], \{A \cap [0,1] \mid A \in \mathcal{B}\})$
- $(\mathbb{R}^n, \mathcal{B}^{(n)}) = (\mathbb{R}, \mathcal{B}) \times \cdots \times (\mathbb{R}, \mathcal{B})$ — $\mathcal{B}^{(n)}$ is the smallest σ-algebra containing all the measurable rectangles $A_1 \times \cdots \times A_n$, where $A_i \in \mathcal{B}$
- Cantor space $\{0,1\}^\omega$, the space of infinite sequences of coin flips, with Borel sets generated by the intervals $\{\alpha \in \{0,1\}^\omega \mid x \leq \alpha\}$ for $x \in \{0,1\}^*$

Examples

Q. Why not always take all subsets 2^S as the measurable sets?

Examples

Q. Why not always take all subsets 2^S as the measurable sets?
A. For $[0, 1]$, it breaks ZFC

Measures

- A measure on $(S, \mathcal{B})$ is a map $\mu : \mathcal{B} \to \mathbb{R}$ such that for any countable collection $\mathcal{A}$ of pairwise disjoint measurable sets, $\mu(\bigcup \mathcal{A}) = \sum_{A \in \mathcal{A}} \mu(A)$ and $\mu(\varnothing) = 0$
- A measure is a probability measure if all $\mu(A) \geq 0$ and $\mu(S) = 1$
- A measure is a subprobability measure if all $\mu(A) \geq 0$ and $\mu(S) \leq 1$
- The measures on $(S, \mathcal{B})$ form a Banach space $\mathcal{M}$ (complete normed vector space) over $\mathbb{R}$
- addition and scalar multiplication are defined pointwise:

$$(a\mu + b\nu)(A) = a\mu(A) + b\nu(A)$$

- the norm is the total variation norm $\|\mu\| = \sup_{A \in \mathcal{B}} |\mu(A)|$

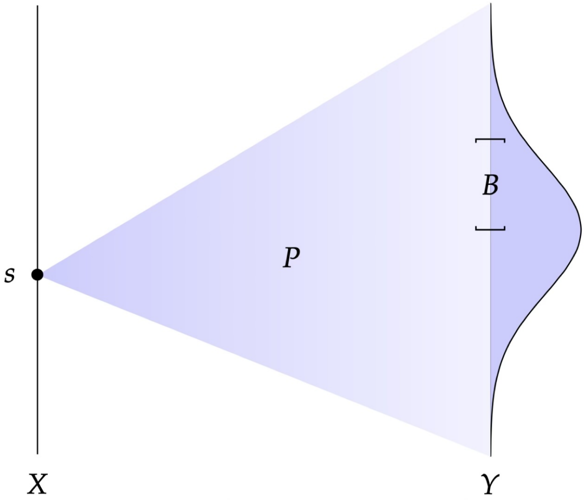
Examples

- Dirac (point mass) measure on a point s:

$$\delta_s(A) \triangleq \begin{cases} 1, & s \in A, \\ 0 & s \notin A \end{cases}$$

- Uniform (Lebesgue) measure λ on $[0,1]$ generated by $\lambda([a,b)) = b - a$
- Uniform (Lebesgue) measure λ on Cantor space $\{0,1\}^\omega$ generated by $\lambda(\{\alpha \mid x \leq \alpha\}) = 2^{-|x|}$

Markov Kernels



Markov Kernels

- A Markov kernel is a function $P : S \times \mathcal{B} \to [0,1]$ such that
  - for fixed $s \in S$, $P(s, -)$ is a subprobability measure $\mathcal{B} \to [0,1]$
  - for fixed $B \in \mathcal{B}$, $P(-, B)$ is a measurable function $S \to [0,1]$
- probabilistic analog of binary relations
- a program p will denote a Markov kernel $[\![p]\!] : S \times \mathcal{B} \to [0,1]$
- $[\![p]\!](s, B)$ is the probability of event B on output, starting in input state s

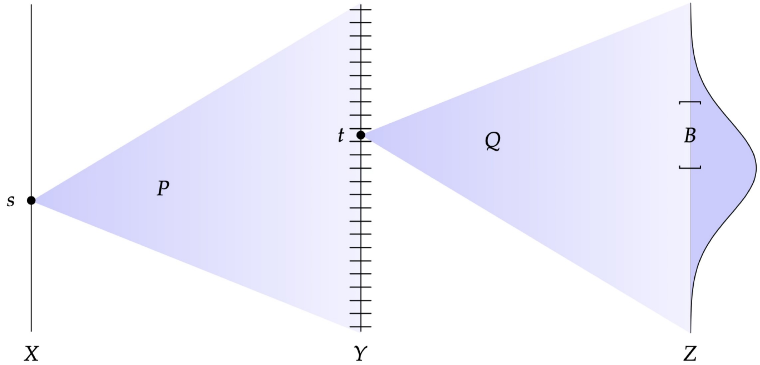
Composition of Markov Kernels

- To compose two Markov kernels, we integrate "up the middle" using Lebesgue integration

$$(P\,;Q)(s,A) \triangleq \int_t P(s,dt)\cdot Q(t,A)$$

- probabilistic analog of relational composition or Boolean matrix multiplication
- just multiplication of stochastic matrices in the discrete case

Composition of Markov Kernels



$$(P;Q)(s,B) \triangleq \int_t P(s,dt) \cdot Q(t,B)$$

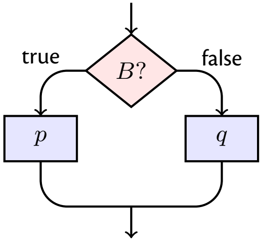
Lifting

- A Markov kernel can be lifted to a linear map $\mathcal{M} \to \mathcal{M}$ via integration

$$P(\mu)(A) \triangleq \int_t \mu(dt) \cdot P(t, A)$$

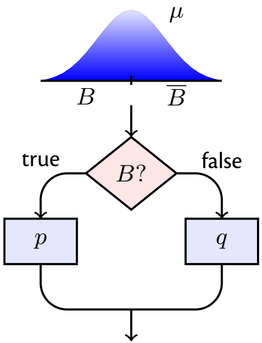
- this map is linear and continuous on $\mathcal{M}$ (since $\int$ is)
- this is Kleisli lifting again

Probabilistic conditional



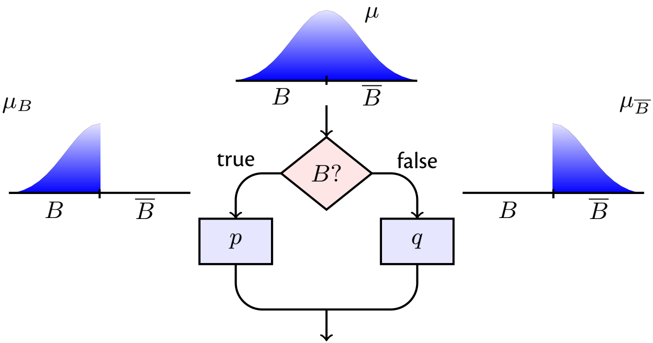
$\texttt{if } B \texttt{ then } p \texttt{ else } q$

Probabilistic conditional



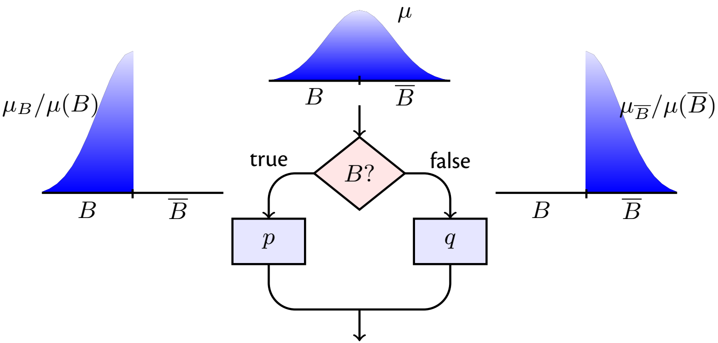
$$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$$

Probabilistic conditional



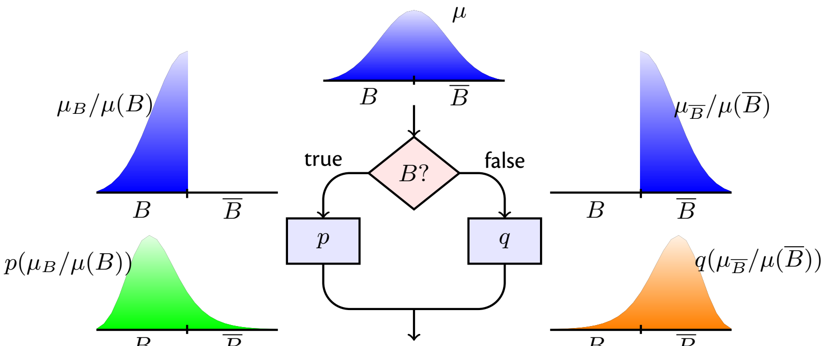
$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$

Probabilistic conditional



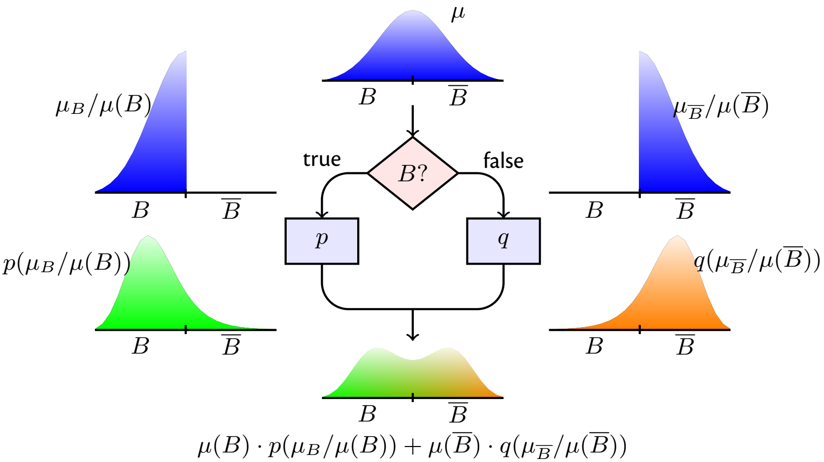
$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$

Probabilistic conditional



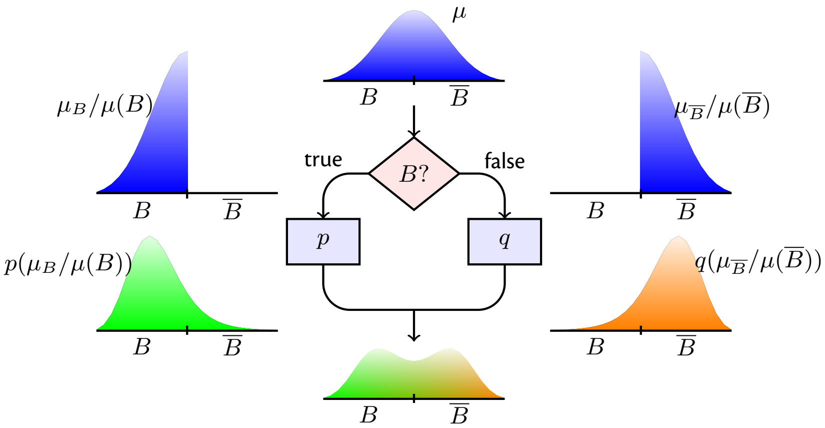
$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$

Probabilistic conditional



$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$

Probabilistic conditional

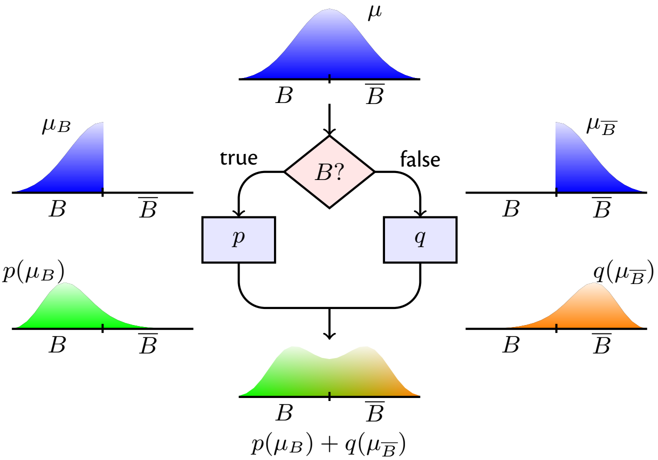


$$\cancel{\mu(B)} \cdot p(\mu_B/\cancel{\mu(B)}) + \cancel{\mu(\overline{B})} \cdot q(\mu_{\overline{B}}/\cancel{\mu(\overline{B})})$$

$$p(\mu_B) + q(\mu_{\overline{B}})$$

`if` B `then` p `else` q

Probabilistic conditional



$\texttt{if}\ B\ \texttt{then}\ p\ \texttt{else}\ q$

Measure Transformer Semantics

- $[\![x := t]\!](s) \triangleq \delta_{s[s(t)/x]}$ (point mass on $s[s(t)/x]$)
- $[\![p\,;q]\!](s) \triangleq [\![q]\!]([\![p]\!](s))$
- $[\![\texttt{if } B \texttt{ then } p \texttt{ else } q]\!](s) \triangleq \begin{cases} [\![p]\!](s), & \text{if } s \in B \\ [\![q]\!](s), & \text{if } s \notin B \end{cases}$
- Define $[\![B]\!](s, A) \triangleq \begin{cases} 1, & s \in A \cap B \\ 0, & s \notin A \cap B \end{cases}$
- Curried and lifted, $[\![B]\!](\mu) = \mu_B$, the measure with all mass outside of B annihilated
- $[\![\texttt{if } B \texttt{ then } p \texttt{ else } q]\!] = [\![B\,;p]\!] + [\![\overline{B}\,;q]\!]$

Measure Transformer Semantics

Random assignment

$$[\![x_i := \mathtt{rnd}]\!](s_1, \ldots, s_n, A_1 \times \cdots \times A_n) \triangleq \Big(\prod_{j \neq i} \delta_{s_j}(A_j)\Big) \cdot \lambda(A_i)$$

where λ is the uniform measure on $[0, 1)$

Measure Transformer Semantics

While loop

$$
\begin{aligned}
[\![\texttt{while}\ B\ \texttt{do}\ p]\!] &\triangleq [\![\overline{B}]\!] + [\![B\,;p\,;\overline{B}]\!] + [\![B\,;p\,;B\,;p\,;\overline{B}]\!] + \cdots \\
&= \sum_n [\![(B\,;p)^n\,;\overline{B}]\!] \\
&= [\![(B\,;p)^*\,;\overline{B}]\!],
\end{aligned}
$$

where

$$
[\![p^*]\!] \triangleq \sum_n [\![p^n]\!]
$$

(although * can produce unbounded measures)

Predicate Transformer Semantics

What can you observe about a state?

- Previously it was whether a state satisfied a predicate: $s \vDash A$
- With probability distributions μ, we can ask about the probability of an event A—this is just $\mu(A)$
- More generally, we can ask about the expected value of a measurable function $f : \mathbb{R}^n \to \mathbb{R}$—this is the integral

$$\int_t \mu(dt) \cdot f(t)$$

- This is the probabilistic analog of $s \vDash A$

Predicate Transformer Semantics

- The function

$$\langle p \rangle f \triangleq \int_t [\![p]\!](-, dt) \cdot f(t)$$

  is a measurable function of the input state—gives expected value of f after executing p
- This is a generalized predicate transformer $\langle p \rangle : \mathcal{F} \to \mathcal{F}$, where $\mathcal{F}$ is the space of bounded measurable functions
- This is also currying and lifting:

$$\mathbb{R}^n \times \mathcal{B}^{(n)} \to [0,1] \overset{\text{curry}}{\cong} \mathcal{B}^{(n)} \to \mathcal{F} \overset{\text{lift}}{\cong} \mathcal{F} \to \mathcal{F}$$

- $\mathcal{F}$ is a Banach space over $\mathbb{R}$ (complete normed linear space), pointwise addition and scalar multiplication, norm $\|f\| = \sup_{\|s\|=1} |f(s)|$

Predicate Transformer Semantics

A logic based on this idea—PPDL

- $\langle p\rangle A$ = probability that A is satisfied on output—probabilistic analog of weakest precondition
- $\langle p\rangle f$ = expected value of random variable f on output
- $[p]A \triangleq 1 - \langle p\rangle(1 - A) = 1 - \langle p\rangle 1 + \langle p\rangle A = [p]0 + \langle p\rangle A$
- $[p]A$ = probability that A is satisfied on output or p doesn't halt—probabilistic analog of weakest liberal precondition
- $A \leq [p]B$ = probabilistic analog of Hoare pca $\{A\}\ p\ \{B\}$

Predicate Transformer Semantics

Axioms of PPDL

$$\langle ap + bq \rangle f = a\langle p \rangle f + b\langle q \rangle f$$
$$\langle p \rangle (af + bg) = a\langle p \rangle f + b\langle p \rangle g$$
$$0 \leq f \Rightarrow 0 \leq \langle p \rangle f$$
$$f \leq g \Rightarrow \langle p \rangle f \leq \langle p \rangle g$$
$$\langle p;q \rangle f = \langle p \rangle \langle q \rangle f$$
$$\langle B \rangle f = Bf$$

$$[p]f = [p]0 + \langle p \rangle f$$
$$0 \leq f \Rightarrow \langle p^* \rangle f = f + \langle p \rangle \langle p^* \rangle f$$
$$0 \leq f \Rightarrow \langle p^* \rangle f = f + \langle p^* \rangle \langle p \rangle f$$
$$0 \leq f + \langle p \rangle g \leq g \Rightarrow \langle p^* \rangle f \leq g$$
$$f \leq 1 \Rightarrow \langle p \rangle f \leq 1$$
$$0 \leq B \leq BB \leq 1$$

Duality

Write

(μ, f)	for	$\int_t \mu(dt) \cdot f(t)$	Lebesgue integral
$\mu\langle p\rangle$	for	$\int_t \mu(dt) \cdot [\![p]\!](t, -)$	measure transformer
$\langle p\rangle f$	for	$\int_t [\![p]\!](-, dt) \cdot f(t)$	predicate transformer

Then

$$(\mu, \langle p\rangle f) = (\mu\langle p\rangle, f)$$

Denotational Semantics

An ω-CPO is a poset $(A, \sqsubseteq)$ such that every countable chain has a supremum

Theorem (Saheb-Djahromi 1980)

The measures on the Borel sets of the Scott topology on an ω-CPO form an ω-CPO.

Thanks!

Our sponsors make OPLSS possible.

Diamond Sponsor

Jane Street

https://www.janestreet.com/

Platinum Sponsors

National Science Foundation
WHERE DISCOVERIES BEGIN

https://www.nsf.gov/

SIGPLAN

https://www.sigplan.org/

Silver Sponsors