Supermartingales, Ranking Functions and Probabilistic Lambda Calculus

Abstract

Our first answer is the notion of *sparse ranking function*. Most of the individual execution steps of a typical program are trivial and easy to mentally skip over. Sparse ranking functions define $S := I^{\mathbb{N}}$, with the Borel -algebra $_S$. Equivalently, a basis of measurable sets is $\prod_{i=0}^{\infty} X_i$ where X_i are all Borel and all but finitely many are I, and the probability measure $_S$ is given by $_S(\prod_{i=0}^{\infty} X_i) := \prod_{i=0}^{\infty} \mathrm{Leb}(X_i)$, writing.96264Tf1001225.833697.168.471697.134Tm[(Leb)]TJ/F1039.96264Tf1001288.40669

$(Y_n)_n$

terminates in 3 steps with probability 1, but isn't rankable because $(Y\ \ x{:}x)\,\underline{0}$ is reachable, although that has probability 0. Not only is this counterexample AST, it's PAST.

V. ANTITONE RANKING FUNCTIONS

Example V.1 (Random walk). (i)

shorthand 2

in, select them accordin to the position in the term where they'll be used instead.

A ***position*** is a finite sequence of steps into a term, defined inductively as

$$::= \quad j$$

Reduction sequences are used rather than reachable skeletons because if the same skeleton is reached twice, different samples may be needed.

Example VI.3. Consider the term $M = Y(\lambda f x.\,\mathrm{if}(\mathrm{sample} - \underline{0.5} < 0, f\,x, x))\,\underline{0}$, which reduces after a few steps to $N = \mathrm{if}(\mathrm{sample} - \underline{0.5} < 0, M, \underline{0})$. If we label samples by just skeletons and positions, and the pre-selected sample for $(Sk(N), \mathrm{if}_1, \underline{\ }_1)$ is less than 0.5, N reduces back to M, then N again, then the same sample is used the next time, therefore it's an infinite loop, whereas if samples are labelled by reduction sequences, the samples for $M \to N$ are independent from the samples for $M \to N \to M \to N$, and so on.

The reduction sequences of skeletons will often be discussed

Before defining the new version of the reduction relation red, the following lemma is necessary for it to be well-defined.

Lemma VI.7.

References

[1] M. O. Rabin, "Probabilistic algorithms," in *Algorithms and complexity: new directions and results*. Academic Press, New York, 1976, pp. 21–39.

[2] A. D. Gordon, T. A. Henzinger, A. V. Nori, and S. K. Rajamani, "Probabilistic programming," in *Proceedings of the on Future of Software Enginee9ing, FOSE 2014, Hyderabad, India, May 31 - June 7, 2014*, 2014, pp. 167–181.

[3] T. Rainforth, "Automating inference, learning, and design using probabilistic programming," Ph.D. dissertation, University of Oxford, 2017.

[4]

[40] O. Kiselyov, “Problems of the lightweight implementation of probabilistic programming,” in *Proceedings of Workshop on Probabilistic Programming Semantics*, 2016.

[41] C.-K. Hur, A. V. Nori, S. K. Rajamani, and S. Sammuel, “A Provably Correct Sampler for Probabilistic Programs,” in *FSTTCS*mming Semantics

$$\frac{}{;\, x : A \;`\; x : A} \qquad \frac{;\, x : A \;`\; M : B}{`\;\; x{:}M : A \;!\; B} \qquad \frac{`\; M : A \;!\; B \qquad `\; N : A}{`\; M\, N : B}$$

is generated from a base type with a -type constant symbol r, and a function symbol $? : A$ for each function type A.)

$$\mathrm{p}\mathrm{sample}\mathrm{q} := r$$

$$\mathrm{p}y\mathrm{q} := y$$

$$\mathrm{p}\ y{:}N\,\mathrm{q} := \ y{:}\mathrm{p}N\,\mathrm{q}$$

$$\mathrm{p}M$$

APPENDIX C
SUPPLEMENTARY MATERIALS FOR SEC. V

Theorem V.4. *Let $(Y_n)_{n \quad 0}$ be an antitone strict supermartingale w.r.t. stopping time T. Then $T < 1$ a.s.*

Proof. First, as

The term $^{n}\underline{x}$ (for $n > 0$, $x \quad 2 < p$) reduces to either $^{n \quad 1}\underline{x+1}$, $^{n}\underline{x+1}$ or $^{n+1}\underline{x+1}$, with probabilities $p \quad x \quad 2$, $(1 \quad p + x \quad 2)\frac{p \quad x \quad 2}{2}$ and $(1 \quad + x \quad 2)(1 \quad \frac{p \quad x \quad 2}{2})$ respectively. Let $a(\ ^{n}\underline{x}) = n + 2x \quad 1$. This is a supermartin-

1

to offset the average increase in n), but it does not satisfy the antitone-strict progress condition. It does however have a bounded-below variance, so the usual method of using $\ln a$ instead of just a works. Calculating the exact amount that $\ln a$ decreases is not necessary, because it can be bounded as follows: a changes by at least 1=2 with probability at least $=$ (assuming $x \quad 3$), therefore $\ln(a)$ decreases in expectation by at least $\frac{1}{11}(\ln(a + \frac{1}{2}) \quad \ln(a) \quad \frac{1}{2a})$, using the fact that a linear approximation to ln, applied to a, at least doesn't increase, then adding the deviation of ln from

$\frac{x \quad a}{a}$

the linear approximation is everywhere an overestimate), we obtain a sufficiently strong bound on the decrease of $\ln(a)$ that it's an antitone sparse ranking function. (It can obviously be extended to

The sequence $(y_n)_{n\ \ 0}$


1
2
2
1

$; @_1; \quad ;$

Figure 3. Illustration of the $_c$ rules.

Lemma D.1. *If* $(N_1; \ _1)$ $(N_2; \ _2)$*, there is some descend-*

$\begin{matrix} 0 & 0 & & 1 & & 2 & & & 0 \\ 1 & 2 & & & & & & & \end{matrix}$

$N \ ; \quad _p \ N \ ; \quad _c \ N \quad _p \ N \ ; \quad .$

$\#$

$) \qquad)$ if $\qquad) \qquad)$

c

$p \qquad\qquad 2 \quad 2$

ants $N \ ; N$ *of* N *resp.* N *, and a position in both of them such that* $(\ _1 \ \ _1)$ $(\ ^0_1 \ \ ^0)$ $(\ ^0_2; \ ^0)$ $(\ _2 \ \ _2)$

Proof. The $_p$ relation can be split into $_\# \ [\ \ _{''}$, where $(A; \ )$ $(B; \ )$ if $A \ ! \ B$ and $(A; \ )$ $_p$ $(B; \ )$, and similarly $(A; \ \ _{''} (B; \ \ B \ ! \ A$ and $(A; \ \ _p (B; \ .$ At each stage of this proof, it will be assumed that the steps and the steps in the sequence from $(N_1; \ _1)$ to $(N \ ; \ )$

The remaining cases are when the subsequence of reduction steps involved in $Y \quad _c\ X$ occur before the last out-of-order pair in X, but possibly overlapping. If there is no overlap, then the $\quad _c$ steps do not interfere with each other at all, and can simply be performed in the other order. This gives a different sequence of $\quad _c$ steps $Y \quad _c\ Y_1 \quad _c\ X_1 \quad _c\ X_{cbv}$. As Y_{cbv} =
1 $(Y\ )_{cbv}$, we can proceed by induction in this case and use the fact that

number of instances of x. Next, each of the images of
to the right of ; ^{0}

part excluded from ^{0}), and the positions of the relevant variable within the skeleton at . They do not depend on or , so that if both initial positions were moved somewhere else the relative positions of the images of

within were swapped with , the relative positions of

in case f, the positions of the relevant variable are taken after the inner reduction takes place, so that some of those may still change). Because there are so many of them and they don't actually affect the multiset of positions where reductions take place, we will be ignoring swaps by case a in the proof of this case, and just assuming that all the

Proof. This is a simple induction on . In the base case, $(A; \ ) = (M; \ )$ therefore $(A; \ ) \ _p (M; \ )$ trivially. Otherwise, suppose that the case that $O|f(O[r]) = \underline{x}(X;\ldots;X)$, a combination of the measurable function x with a rearrangement of vector components. In the case that $O \in R_S$, the function $_O$ simply inserts the sample into the vector of reals at a certain index, call it h

Theorem VI.10. *A closed term M is AST with respect to* cbv *iff it is AST.*

Proof. For both red and $)_{\text{cbv}}$

A suitable sparse ranking function for it is

$$M \mapsto \frac{2}{3}$$

$$P[\underline{n}] \mapsto 2n + 2$$

$$(\ p{:}\ \ [p]\ \underline{n})\ (\ x{:}\ \underbrace{x \ \cdots\ x}_{n\ \ k}\ (\ \ [x]\ \underline{k})) \mapsto n + k + 2$$

$$\underbrace{\qquad\qquad}_{n}$$