

S-finite Kernels and Game Semantics for Probabilistic Programming

Luke Ong Matthijs Vákár
University of Oxford

1 Introduction

It is a well-known result that higher-order functions cannot generally be interpreted within the confines of standard measure theory, in the sense that the category of measurable spaces and measurable functions is not cartesian closed. This has led authors to consider more general semantic structures, like quasi Borel spaces, for analysing continuous distribution probabilistic programs involving higher-order functions [3]. In this paper, we argue that this lack of measurable space structure on sets of measurable functions is due to the extensional nature of the homset and that classical measure theory is entirely sufficient for studying higher-order probabilistic computation if we work with a more intensional notion of function. In particular, we present a framework for “measurable game semantics” as a natural combination of traditional game semantics with standard measure theory.

Staton [7] has argued convincingly that s-finite kernels, a little studied class of kernels, are necessary and sufficient for the semantics of first-order probabilistic programs. In our game semantics, we show how they can also be used to give semantics for higher-order probabilistic programs. For this reason, we first establish some of their basic theory.

As a proof of concept, we demonstrate that our measurable game semantics suffices to give a model of a higher-order probabilistic programming language with continuous distributions. A strength of game semantics is that we can choose to include or exclude various classes of effects like recursion, local state and control operators by varying our notion of strategy in order to obtain fully abstract models for the corresponding languages. To demonstrate the flexibility of our framework, we present the first model (to our knowledge) of a probabilistic programming language with continuous distributions, higher-order functions, recursion and local mutable variables of type \mathbb{R} . We show full abstraction of our model for a call-by-name operational semantics.

A pleasant by-product of our approach is a canonical construction of standard Borel spaces of program traces, on which inference algorithms like trace MCMC and SMC can operate, as spaces of plays of the games that types denote.

2 A Higher-Order Imperative Probabilistic Language (PIA)

We use *Probabilistic Idealised Algol* (PIA), an extension of Idealised Algol [6] with a ground type \mathbb{R} of real numbers and probabilistic programming constructs. To be precise, we consider types

$$T ::= \mathbb{R} \mid \text{Var}[\mathbb{R}] \mid \text{void} \mid T \Rightarrow T.$$

In addition to the usual constructs of the simply-typed λ -calculus with fixpoint combinators Y at all types, we consider, in order to implement references of type \mathbb{R} , for $W : \text{Var}[\mathbb{R}]$ and $V : \mathbb{R}$, a command $(W := V) : \text{void}$, a value $\text{deref}(W) : \mathbb{R}$ and an operator $\text{new}_{x:\text{Var}[\mathbb{R}]}$ binding a variable $x : \text{Var}[\mathbb{R}]$. (Additionally, we include the usual primitive mkvar .) We add primitives for all measurable functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (including constants $r : \mathbb{R}$). We further consider the following two probabilistic constructs: a primitive $\text{sample} : \mathbb{R}$ for sampling from the uniform distribution on $[0, 1]$ and a command $\text{score} : \mathbb{R} \Rightarrow \text{void}$ for reweighting program traces in order to implement soft constraints.

Let us use the syntactic sugar $\text{let } x \text{ be } M \text{ in } N$ for $\text{new}_{y:\text{Var}[\mathbb{R}]}(y := M ; N[\text{deref}(y)/x])$, where y is fresh. This gives us a CBV substitution for terms of ground type which is useful for copying the results of effectful computations. By the Randomisation Lemma [5], all probability kernels from \mathbb{R}^n to \mathbb{R} are definable in PIA as $f(\text{sample})$ for some measurable function $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$. Thus we add them as syntactic sugar as well. As usual, we can define while loops using conditionals (which are included as a measurable function) and recursion.

As an example of the expressive power of PIA, note that we can represent the beta-binomial distribution $\text{betabinom}(n, \alpha, \beta)$ in (at least) two ways: purely functionally, using a (continuous) beta prior for a binomial distribution,

$$\text{binom}(n, \text{beta}(\alpha, \beta)),$$

and imperatively, using a (discrete) Pólya urn scheme

$$\text{new}_{k,l:\text{Var}[\mathbb{R}]} \left(k := 0 ; l := 0 ; \text{while}(k < n) \left(k := \text{deref}(k) + 1 ; \right. \right. \\ \left. \left. l := \text{deref}(l) + \text{flip}\left(\frac{\alpha + \text{deref}(l)}{\alpha + \beta + \text{deref}(k) - 1}\right) \right) ; \text{deref}(l) \right).$$

3 Standard Borel Spaces

We recall that standard Borel spaces can be defined to be measurable spaces that are isomorphic to either a countable space or \mathbb{R} . They provide a convenient setting for measure theory. For our purposes, the important facts about these are that they are closed under countable products, coproducts and measurable subspaces. Other useful observations are that a function $f : X \rightarrow Y$ between such spaces is measurable iff its graph is a measurable subset of $X \times Y$ and that a measurable injection between such spaces is an embedding.

4 S-finite Measures and Kernels

A measure is called s-finite if it is a countable sum of finite measures. Similarly, a kernel is called s-finite if it is a countable sum of finite kernels (kernels with a uniform finite bound). Equivalently, a kernel is s-finite iff it is the pushforward of a σ -finite kernel (a kernel k which is the countable sum of kernels $(k_i)_{i \in \mathbb{N}}$ such that k_i is supported in Y_i for some fixed measurable partition $\bigcup_{i \in \mathbb{N}} Y_i$ of the codomain of k). It is an important fact that s-finite measures are closed under pushforward and that s-finite kernels are closed under composition (kernel integration) [7]. Moreover, because they satisfy a unique extension theorem, we can observe that Fubini's Theorem for swapping order of integration holds for s-finite measures. [7] has argued convincingly that s-finite kernels are necessary and sufficient for giving semantics to first-order probabilistic programs.

We have the following characterisations of s-finite measures and kernels.

Theorem 4.1. *A measure ν on X is s-finite iff there exists a σ -finite measure μ on X and a measurable function $f : X \rightarrow \{1, \infty\}$ such that ν has density f w.r.t. μ iff there exists a probability measure μ on X and a measurable function $f : X \rightarrow [0, \infty]$ such that ν has density f w.r.t. μ (ν is σ -finite iff this density can be chosen to be finite).*

Conversely, an s-finite measure ν is σ -finite iff there exists a measurable $f : X \rightarrow (0, \infty)$ such that $\int_X \nu(dx) f(x) < \infty$.

Theorem 4.2 (Randomisation). *Let X, Y be standard Borel spaces. Writing $\#$ and \mathbb{U} for the counting and Lebesgue measures, respectively, a kernel k from X to Y is s-finite iff it is the pushforward $\lambda x. f(x, -)_* (\#_{\mathbb{N}} \otimes \mathbb{U}_{[0,1]})$, for some measurable partial function $f : X \times \mathbb{R} \times \mathbb{R} \rightarrow Y$, iff it is the pushforward $\lambda x. f(x, -)_* \mathbb{U}_{[0,\infty)}$ for some measurable partial function $f : X \times \mathbb{R} \rightarrow Y$.*

Let us call a measurable subset U of X an ∞ -set w.r.t. μ if $\mu(U) = \infty$, and $\mu(V) = 0$ or ∞ for all measurable $V \subseteq U$.

Theorem 4.3. (i) σ -finite measures have no ∞ -sets.
 (ii) An s-finite measure μ which is not σ -finite has a unique (up to μ -null sets) maximal ∞ -set, written $\infty[\mu]$, outside of which it is σ -finite.

This lets us formulate a Radon-Nikodým Theorem for s-finite measures.

Theorem 4.4 (Radon-Nikodým). *Let μ, ν be s-finite measures on X . Then, μ has a density (or Radon-Nikodým derivative) $f : X \rightarrow [0, \infty]$ w.r.t. ν iff $\mu \ll \nu$ and $\mu(\infty[\nu] \setminus \infty[\mu]) = 0$.*

This is relevant for probabilistic programming, for instance, as it gives the precise conditions under which the following importance sampling procedure is valid:

$$\mu = \text{let } x \text{ be } \nu \text{ in score}(f(x)); x,$$

as [7] has argued that probabilistic programs of ground type represent s-finite measures. In particular, as observed by [7], we can define $\#_{\mathbb{N}}$ as

$$\#_{\mathbb{N}} = \text{let } x \text{ be poisson}(1) \text{ in score}(x! \cdot e); x,$$

meaning that all s-finite kernels become definable in PIA by randomisation.

Similarly, s-finite measures have a well-behaved theory of conditional probability in the form a Disintegration Theorem.

Theorem 4.5 (Disintegration). *Let X, Y be standard Borel spaces, let $\phi : X \rightarrow Y$ measurable and let μ and ν be s-finite measures on X and Y respectively. Then, μ has an s-finite (kernel) disintegration k w.r.t. ν iff $\phi_* \mu \ll \nu$ and $\mu(\phi^{-1}(\infty[\nu]) \setminus \infty[\mu]) = 0$. Recall that we call a kernel k a disintegration iff $\nu; k = \mu$ and $k(x)$ is supported in $\phi^{-1}(x)$ for ν -almost all x .*

We can formulate uniqueness properties for Radon-Nikodým derivatives and disintegrations but have to be a bit careful on ∞ -sets where non-zero rescalings are possible.

5 Measurable Game Semantics

The idea of game semantics is to model a computation as a play in a turn-based 2-player game between a program (Player) and its environment (Opponent), where Opponent gets to move first. In this view, a program becomes a strategy on a game specified by its type. As we will explain, to construct a semantics for the types of PIA, we work with standard HO-games A [2, 4], where we equip the set of moves (tokens) M_A with a standard Borel space structure, and require the enabler and question-answer labelling functions to be measurable. We call the resulting structure a *measurable game*. We note that this together with the countable (co)product closure of standard Borel spaces means that the set of plays \mathcal{L}_A (certain sequences of moves) has a standard Borel space structure, to be thought of as a space of program traces of the type A .

We can note that any standard Borel space X gives rise to a measurable game X_* which has a single initial Opponent question $*$, to which Player can respond with any element $x \in X$ as an answer. The interpretation of the type constructors of Idealized Algol follows the standard game semantics recipes, making the crucial observation that in order to construct $A \times B$ and $A \Rightarrow B$, it suffices to take finite products and coproducts of the spaces of moves of A and B , which exist as standard Borel spaces. In particular, due to the intensional

nature of the game-semantic function space, we circumvent the usual issues of constructing \Rightarrow -types in a continuous probability setting. `void` is interpreted as the game with a single initial Opponent question run, followed by a single Player answer done.

We can define a *deterministic strategy (in global form)* on a measurable game A as one usually would, as a non-empty prefix-closed set μ of even-length plays $\mathcal{L}_A^{\text{even}}$ such that for odd-length plays s there is at most one extension $sa \in \mu$ of s by one move, but with the extra conditions that μ and the domain of μ (the odd-length prefixes of μ) are measurable subsets of \mathcal{L}_A . We will refer to a subclass of these strategies, called *innocent strategies*, which are only allowed to base their response on part of the history of the play so far. A more intensional description of strategies is given by strategies in *local form*: measurable partial functions $\sigma : \mathcal{L}_A^{\text{odd}} \rightarrow \mathcal{L}_A^{\text{even}}$ such that $\sigma(s)$ is \perp or an extension of s by one move. We note that every local form defines a global form and that a global form has at least one local form.

In particular, we have an (innocent) *copycat* (or identity) strategy on $A \Rightarrow A$ which always copies the last move from one copy of A to the other. Given deterministic strategies on $A \Rightarrow B$ and $B \Rightarrow C$ we can compose them using their interaction, as usual. The fact that this preserves measurability can be proved using the fact that measurable injections between standard Borel spaces are embeddings. This gives us a cartesian closed category of measurable games and deterministic strategies (in global form). In fact, this setting suffices to interpret the fragment of PIA with the exception of the constructs `sample` and `score`, noting that measurable strategies are closed under suprema of ω -chains because the countable union of measurable sets is measurable to interpret `Y` and following the recipes of [1] for interpreting the imperative constructs.

To interpret `sample` and `score`, we would like to generalise to a suitable notion of *weighted strategy*. It is perhaps most clear what a good notion of weighted strategy in local form should be: an s-finite kernel $\sigma : \mathcal{L}_A^{\text{odd}} \rightsquigarrow \mathcal{L}_A^{\text{even}}$ such that $\sigma(s)$ is supported in the extensions of s with one move. We call such a strategy a *subprobability strategy* if it is in fact a subprobability kernel. To see what a good notion of composition is for such σ , note that we can use randomisation to obtain an even more intensional description of σ as $\lambda s.f(-, s) * \bigotimes_{[0, \infty)}^{\mathbb{N}} \mathbb{U}_{[0, \infty)}$ for a measurable partial function $f : \mathbb{R}^{\mathbb{N}} \times \mathcal{L}_A^{\text{odd}} \rightarrow \mathcal{L}_A^{\text{even}}$ such that for all r , $f(r, -)$ is a deterministic strategy in local form where $r(n)$ only affects the response of $f(r, -)$ at plays of length n .

We have copycat strategies in such form, which, for $r \in [0, 1]$, act as copycats and otherwise diverge. We can compose such weighted strategies in randomised form f and g by

$$f ; g := \lambda r.f(\Pi_{\mathbb{N}}\text{odd}(r)/2) ; g(\Pi_{\mathbb{N}}\text{even}(r)/2),$$

where we use the usual composition of deterministic strategies and where odd and even expand a real number in binary

and take the odd and even digits, respectively. The point is that $\text{even}/2 * \mathbb{U}_{[0, \infty)} = \text{odd}/2 * \mathbb{U}_{[0, \infty)} = \mathbb{U}_{[0, \infty)}$, thus ensuring that f and g have access to their own independent source of randomness.

Next, we define a good notion of weighted strategy in global form μ on A associated with a local form σ as a certain kind of measure on the complete even length plays in A . The intuition is that μ is defined via the Carathéodory Extension Theorem to be the maximal measure μ which, on any deterministic Opponent strategy on A , will restrict to the parallel composition of σ and τ ; and μ will assign measure ∞ to any set which cannot be covered by countably many deterministic Opponents. This parallel composition can be defined either by choosing a randomised form for σ and τ and using the parallel composition of randomised forms, after which we feed in the oracle $\bigotimes_{[0, \infty)}^{\mathbb{N}} \mathbb{U}_{[0, \infty)}$; or it can be defined directly using σ and τ by an iterated integral.

We note that the quotient from weighted strategies in randomised form to global form defines a congruence relation - essentially because Fubini's Theorem holds for s-finite kernels - and that we obtain a rational cartesian closed category with the category of deterministic strategies as a lluf subcategory. Moreover, this category has a natural interpretation of `sample` and `score`. The former is interpreted by the weighted strategy on \mathbb{R}_* with local form which responds with $\mathbb{U}_{[0, 1]}$ to the unique initial move $*$. The latter is interpreted by the weighted strategy on $\mathbb{R}_* \Rightarrow \text{void}$ which responds to the initial move `run` with the initial move $*$ in \mathbb{R}_* and to the move $r \in \mathbb{R}$ with the measure $|r| \cdot \delta_{\text{done}}$.

We note that our semantics equates our two descriptions of a beta-binomial distribution.

6 Definability

One can prove that all (basis¹) weighted strategies at our type hierarchy in the model are definable in PIA. To do this, one notes that a weighted strategy μ on A can be factored as $\#_{\mathbb{N}}; \text{sprob}(\mu)$ for a subprobability strategy $\text{sprob}(\mu)$ on $\mathbb{R} \Rightarrow A$. (And, we have already seen that $\#_{\mathbb{N}}$ is definable in PIA.) Further, we can note, by the usual Randomisation Lemma for subprobability kernels between standard Borel spaces, that any subprobability strategy μ on A can be factored as `sample`; $\text{det}(\mu)$ where $\text{det}(\mu)$ is a deterministic strategy on $\mathbb{R} \Rightarrow A$. Next, as usual, any deterministic strategy μ on A factors as $\text{new}_{\text{Var}[\mathbb{R}]}$ ($\text{inn}(\mu)$), where $\text{inn}(\mu)$ is a deterministic innocent strategy on $\text{Var}[\mathbb{R}] \Rightarrow A$ which essentially writes and reads the whole history of the play into/from the variable of type \mathbb{R} . Then, using standard game semantics techniques (known as the Decomposition Lemma) we can define such strategies from deterministic innocent strategies on $\mathbb{R}_*^n \Rightarrow \mathbb{R}_*$ which examine all arguments once (from left to right) or are non-strict, which are definable as $\lambda_x.f(x)$ for a

¹That is, strategies which are supported in plays of length at most N for some $N \in \mathbb{N}$.

measurable partial function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (which is definable from measurable $f' : \mathbb{R}^n \rightarrow \mathbb{R}$ and Y) or definable as $\lambda_x r$ for a constant $r \in \mathbb{R}$. For the usual reasons, this definability result implies full abstraction of the model with respect to call-by-name evaluation, if we equate two strategies if the restriction of their global form to complete plays agrees.

7 Operational Semantics and Full Abstraction

We can define a (call-by-name) small-step operational semantics for PIA as a finite kernel if we equip the syntax with a measurable space structure as described in, for instance, [8]. We can note that this induces a big-step semantics which is an s-finite kernel, as a supremum of an ω -chain of finite kernels. We note that our denotational semantics is adequate with respect to this operational semantics. Our definability result now implies for the usual reasons that our model is fully abstract. Note that this in particular implies that deterministic program contexts of type `void` determine observational equivalence entirely.

8 Future Directions

We would like to find a more analytic characterisation of which measures on complete plays are weighted strategies in global forms. Similarly, we would like to find a more explicit description of composition directly defined on global forms, without relying on randomised forms or local forms or prove that such a description is impossible.

If we consider functions up to $\mathbb{U}_{\mathbb{R}}$ -almost everywhere equality, we suspect we can obtain a notion of global form as an s-finite measure which gives a fully abstract model. The Disintegration Theorem then allows us to construct local forms from local forms. Moreover, it suffices to work with countably many functions in language, using approximation by step-functions.

Relying on standard techniques from the field of game semantics, our results should extend to a broader class of languages. By varying the games and strategies we consider in the usual ways, we believe we can extend our type system with tuples, variant types and recursive types, we can change the evaluation strategy to call-by-value and we can add higher-order references and non-local control features to the language (or remove the ground references we have so far considered).

Acknowledgements. We are thankful to Mario Lezcano for comments on this work. Supported by the Engineering and Physical Sciences Research Council grant EP/M023974/1.

References

- [1] Samson Abramsky and Guy McCusker. 1997. Linearity, sharing and state: a fully abstract game semantics for Idealized Algol with active expressions. In *Algol-like languages*. Springer, 297–329.
- [2] V Danos and R Harmer. 2000. Probabilistic game semantics. In *Logic in Computer Science, 2000. Proceedings. 15th Annual IEEE Symposium on*. IEEE, 204–213.
- [3] Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. 2017. A Convenient Category for Higher-Order Probability Theory. *arXiv preprint arXiv:1701.02547* (2017).
- [4] J Martin E Hyland and C-H Luke Ong. 2000. On full abstraction for PCF: I, II, and III. *Information and computation* 163, 2 (2000), 285–408.
- [5] Olav Kallenberg. 2006. *Foundations of modern probability*. Springer Science & Business Media.
- [6] John C Reynolds. 1997. The essence of Algol. In *ALGOL-like Languages*. Springer, 67–88.
- [7] Sam Staton. 2017. Commutative semantics for probabilistic programming. In *European Symposium on Programming*. Springer, 855–879.
- [8] Sam Staton, Hongseok Yang, Frank Wood, Chris Heunen, and Ohad Kammar. 2016. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*. ACM, 525–534.