S-finite Kernels and Game Semantics for Probabilistic Programming
Luke Ong and Matthijs Vákár
University of Oxford

Summary
- **S-finite kernels are necessary and sufficient** for first-order PPL semantics (Staton).
- However, little studied.
- We establish basic results, relating them to probability and σ-finite kernels.
- establish a Radon-Nikodým theorem, explaining general importance sampling.
- and a disintegration theorem, explaining conditioning in generality.
- We introduce PIA, an idealised higher-order imperative PPL.
- We show that traditional *game semantics* + standard Borel spaces + s-finite kernels $\sim$ fully abstract model of PIA.
- Shows deterministc contexts of type void determine observational equivalence.

Status Quo and Problems
- **S-finite kernels** on measurable spaces are precisely right for modeling first-order functional PPLs, in the sense that we have a definability result (Staton '17).
- Problem 1: poorly understood class of kernels.
- Problem 2: well-known that measure theory cannot accommodate higher-order functions (Aumann '61).
- Fix for 2: quasi-Borel spaces (Hennion et al. '17), more exctic structures that do combine probability and higher-order functions.
- Problem 3: s-finite kernels on these not known to satisfy definability or full abstraction.
- Problem 4: Related paradox: no problems in implementing higher-order PPL.
- Problem 5: quasi-Borel spaces so far have only been used to model functional PPLs, but imperative behaviour is important in practice (e.g. in Stan).

Our Solutions
- Fix for 4: We use a more intensional notion of semantics than quasi-Borel spaces, in the form of game semantics, where standard measure theory does suffice to model higher-order functions.
- Fix for 5: Using standard game semantics techniques, we can model range of effects.
- Fix for 3: PIA definability and full abstraction.

Theory of S-finite Measures and Kernels

**Definition** (Measure/kernel). s-finite if $\Sigma_{\alpha \in \mathbb{N}}$ of (uniformly) finite ones.

**Theorem** (Staton '17) Closed under kernel integration. Equivalent def.: pushforward of σ-finite ones. Satisfy Fubini, so can swap order of integration. Model + definability for 1st-order functional PPL.

We establish the following two characterisations for s-finite measures and kernels.

**Theorem** (Characterising s-finite measures). A measure $\nu$ on $X$ is s-finite iff there exists a σ-finite measure $\mu$ on $X$ and a measurable function $f : X \to [0, \infty]$ such that $\nu$ has density $f$ w.r.t. $\mu$ iff there exists a probability measure $\mu$ on $X$ and a measurable function $f : X \to [0, \infty]$ such that $\nu$ has density $f$ w.r.t. $\mu$ (s-finite iff $f$ is finite).

**Theorem** (Randomisation). Let $X,Y$ be standard Borel spaces. Kernel $k : X \to Y$ is s-finite iff $k = \lambda x.f(x),_{\Sigma_{\alpha \in \mathbb{N}}}$ of the Lebesgue measure for some measurable partial function $f : X \times \mathbb{R} \to Y$.

**Definition** (∞-set). To formulate RN- and disintegration theorems: call $U \subseteq X$ an ∞-set w.r.t. measure $\mu$ if $\mu(U) = \infty$ and $\mu(V) = 0, \infty$ for all $V \subseteq U$.

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

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

Relevant for probabilistic programming - importance sampling: $\mu = \text{let } x \text{ be } v \text{ in } \text{score}(f(x)) \times x$.

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

So well-defined theory of conditional probability.

Can formulate uniqueness properties for Radon-Nikodým derivatives and disintegrations with care on ∞-sets.

Measurable Game Semantics

- **Game semantics**: computation trace as a *play* in a turn-based 2-player game between program (Player) and environment (Opponent).
- **Program - strategy** on game specified by type.
- Framework for giving uniform *fully abstract* semantics for range of effectful languages.
- Flexibly vary type systems and effects, by varying classes of games and strategies.
- Idea 1: as usual + standard Borel space (sbs) structure on moves + “measurable rules” ~ sbs structure on plays/traces. Works because function space small enough s.t. sbs.
- Idea 2: deterministic strategies as measurable partial functions from odd to even plays.
- Idea 3: weighted strategies as above but s-finite kernels from odd to even plays.
- Idea 4: randomisation - define weighted strategies from deterministic, w/ sample, score.
- Idea 5: existing definability results do the rest.

Probabilistic Idealised Algol
(PIA)

- Types: $T := \mathbb{R} | \text{Var}[\mathbb{R}] | \text{void} | T \Rightarrow T$.
- Simple A-calculus with fixpoint combinators $Y$ at each type.
- Algol-style ground references $\text{Var}[\mathbb{R}]$ of type $\mathbb{R}$, with block structure.
- sample for drawing from $U[p]$ and score for reweighting program traces.
- All measurable functions $\mathbb{R}^n \to \mathbb{R}, n \in \mathbb{N}$.
- Randomisation: s-finite kern. as synt. sugar.
- CBN evaluation.
- Synt. sugar CBV binding let $x$ be $M$ in $N = \text{new}_{\text{Var}}[g/y : M[N[deref(g/y)]x]]$.
- E.g. express beta-binomial continuous functional or discrete inverse. Prove =.