Radon-Nikodým derivatives and disintegration for s-finite measures:

some semantic bases for probabilistic metaprogramming

Luke Ong Matthijs Vákár

University of Oxford

Seminar 113: Metaprogramming for Statistical Machine Learning

21-25 May 2018, Shonan Village Centre

Outline

A typed 1st-order probabilistic programming language, PPL

Idealised, 1st-order version of Church, Anglican, Venture, Hakura, etc. (Staton et al. LICS 2016)

PPL Types. $A, B ::= \mathbb{R} \mid \mathsf{P}(A) \mid 1 \mid A \times B \mid \sum_{i \in I} A_i$, where I is countable, nonempty.

- Types A are interpreted as measurable spaces $[\![A]\!]$.
- $[\![\mathbb{R}]\!]$ is the measurable space of reals with its Borel sets.
- $[\![\mathsf{P}(A)]\!]$ is the measurable space of probabilistic measures on $[\![A]\!]$ (i.e. "Giry monad").
- The type of booleans and natural numbers are definable.

PPL Terms-in-context. Two typing judgements:

- $\Gamma \vdash_{\mathsf{d}} t : A$ for deterministic terms
- $\Gamma \vdash_{\mathsf{p}} t : A$ for probabilistic terms

Terms-in-contexts of PPL

Sums and products. The language includes variables, and standard constructors and destructors for sum and product types.

Sequencing: monadic unit, and bind

$$\frac{\Gamma \vdash_{\mathsf{d}} t : A}{\Gamma \vdash_{\mathsf{p}} \mathsf{return}\,(t) : A} \qquad \frac{\Gamma \vdash_{\mathsf{p}} t : A \qquad \Gamma, x : A \vdash_{\mathsf{p}} u : B}{\Gamma \vdash_{\mathsf{p}} \mathsf{let}\, x = t \,\mathsf{in}\, u : B}$$

Language-specific constructs. Constants for all measurable functions.

Bayesian constructs: Posterior $\propto$ Likelihood $\times$ Prior

1. Sampling from prior distributions: $\dfrac{\Gamma \vdash_{\mathsf{d}} t : \mathsf{P}(A)}{\Gamma \vdash_{\mathsf{p}} \mathsf{sample}(t) : A}$
2. Recoding likelihood scores: $\dfrac{\Gamma \vdash_{\mathsf{d}} t : \mathbb{R}}{\Gamma \vdash_{\mathsf{p}} \mathsf{score}(t) : 1}$
3. Normalisation (for posterior): $\dfrac{\Gamma \vdash_{\mathsf{p}} t : A}{\Gamma \vdash_{\mathsf{d}} \mathsf{normalise}(t) : \mathbb{R} \times \mathsf{P}(A) + 1 + 1}$

Semantics of PPL (Staton, ESOP 2017)

- Interpret $\Gamma \vdash_{\mathsf{d}} t : A$ as a measurable function $[\![t]\!] : [\![\Gamma]\!] \to [\![A]\!]$
- Interpret $\Gamma \vdash_{\mathsf{p}} t : A$ as an s-finite kernel $[\![t]\!] : [\![\Gamma]\!] \rightsquigarrow [\![A]\!]$.

DEF. A kernel k from (X, Σ_X) to (Y, Σ_Y) is function $k : X \times \Sigma_Y \to [0, \infty]$ s.t.

i. $\forall x \in X$, $k(x, -) : \Sigma_Y \to [0, \infty]$ is a measure

ii. $\forall U \in \Sigma_Y$, $k(-, U) : X \to [0, \infty]$ is a measurable function.

(Henceforth identify measures with kernels $\mu : 1 \times \Sigma_Y \to [0, \infty]$)

Kernel k(-, -)	Definition
subprobability	$\sup_{x \in X} k(x, Y) \leq 1$
finite	$\sup_{x \in X} k(x, Y) < \infty$
σ-finite	$\exists (Y_i \in \Sigma_Y)_{i \in \omega}.\big(Y = \biguplus_i Y_i \ \& \ \forall i \,.\, \sup_{x \in X} k(x, Y_i) < \infty\big)$
s-finite	$k = \sum_{i \in \omega} k_i$, each k_i is a finite kernel $X \rightsquigarrow Y$.

The classes above form an increasing chain (ordered by $\subseteq$).

Examples of σ-finite / s-finite measures

DEF. Let (X, Σ_X) be a measurable space; $\mu : \Sigma_X \to [0, \infty]$ be a measure.

- μ is σ-finite if $X = \biguplus_{i \in \omega} X_i$ with each $X_i \in \Sigma_X$ and $\mu(X_i) < \infty$.
- μ is s-finite if $\mu = \sum_{i \in \omega} \mu_i$, and each $\mu_i(X) < \infty$.

Intuition: "bad ∞" is ∞ concentrated at a point.

- σ-finiteness only admits "good ∞"
- s-finiteness can admit "bad ∞", but only countably many.

Examples

1. The Lebesgue measure, Leb, is σ-finite.
2. The ∞-measure on the point 1 is s-finite, but not σ-finite;
3. Counting measure $\#_S$ on any uncountable standard Borel space S is not s-finite
4. $\infty \cdot Leb$ is not s-finite. (Convention: $0 \cdot \infty = 0$.)

s-finite-measure semantics of PPL (Staton, ESOP 2017)

Context $\Gamma = (x_1 : A_1, \cdots, x_n : A_n)$, with $[\![\Gamma]\!] := \prod_{i=1}^{n} [\![A_i]\!]$.

Semantics of PPL

- Interpret $\Gamma \vdash_{\mathsf{d}} t : A$ as a measurable function $[\![t]\!] : [\![\Gamma]\!] \to [\![A]\!]$
- Interpret $\Gamma \vdash_{\mathsf{p}} t : A$ as an s-finite kernel $[\![t]\!] : [\![\Gamma]\!] \rightsquigarrow [\![A]\!]$.

Theorem (Definability)

If kernel $k : [\![\Gamma]\!] \rightsquigarrow [\![A]\!]$ is s-finite, then there is a term $\Gamma \vdash_{\mathsf{p}} t : A$ s.t. $k = [\![t]\!]$.

This is a very useful result (for us)!

Why s-finite (and not σ-finite) measures?

Infinite measures seem unavoidable.

- No known useful syntactic restriction that enforces finite measures.
- A program with finite measure may have subexpression with infinite measure.

σ-finite measures are a much-studied class of infinite measures, but they are not suitable for interpreting probabilistic programming languages.

- The pushforward of a s-finite measure is s-finite; but the pushforward of a σ-finite measure is generally only s-finite.
- Failure of kernel composition of σ-finite measures: let $U \in \Sigma_1$

$$[\![\vdash_{\mathsf{p}} \mathsf{let}\, x = \mathit{Leb}\, \mathsf{in\, return}\ () : 1]\!](U) = \int_{\mathbb{R}} \mathit{Leb}(\mathrm{d}z)\, \chi_{()}(U) = \infty \cdot \chi_{()}(U).$$

Leb (Lebesgue measure) is σ-finite, however the composite is s-finite, and not σ-finite

Talk outline

1. Introduction: s-finite-measure semantics of an idealised 1st-order probabilistic programming language
2. Properties of s-finite measures and kernels
3. Radon-Nikodým derivatives
4. Conditional distribution and disintegration
5. Conclusions and further directions

Outline

Product measures

Given measure μ on X and kernel ν from X to Y, call measure Ψ on $X \times Y$ a product measure of μ and ν if $\Psi(U \times V) = \int_U \mu(\mathrm{d}x)\, \nu(x, V)$, for all $U \times V \in \Sigma_{X \times Y}$.

- By Carathéodory Extension Theorem, a maximal product measure $\mu \otimes \nu$ always exists:

$$(\mu \otimes \nu)(W) := \inf \left\{ \sum_{i \in \omega} \int_{U_i} \mu(\mathrm{d}x) \nu(x, V_i) \mid W \subseteq \bigcup_{i \in \omega} (U_i \times V_i) \in \Sigma_{X \times Y} \right\}.$$

- Product measures may be defined via iterated integration:

$$(\mu \otimes^l \nu)(W) := \int_X \mu(\mathrm{d}x) \int_Y \nu(x, \mathrm{d}y)\; \chi_W(x, y)$$

and, in case $\nu(x)$ is independent of x, i.e., ν is a measure on Y

$$(\mu \otimes^r \nu)(W) := \int_Y \nu(\mathrm{d}y) \int_X \mu(\mathrm{d}x)\; \chi_W(x, y).$$

Fubini theorem—for swapping order of integration

Even when $\otimes^l$ and $\otimes^r$ are well-defined, they may not be equal:

- For non-Leb-null $V \in \Sigma_{\mathbb{R}}$:

$$\otimes^l : \quad \int \#_{\mathbb{R}}(\mathrm{d}x)\Big(\int Leb(\mathrm{d}y)\ \{(r,r) \mid r \in V\}\Big) = 0$$

$$\otimes^r : \quad \int Leb(\mathrm{d}y)\Big(\int \#_{\mathbb{R}}(\mathrm{d}x)\ \{(r,r) \mid r \in V\}\Big) = Leb(V).$$

Theorem (Fubini)

For an s-finite measure μ on X, and an s-finite kernel ν from X to Y

i. *$\mu \otimes^l \nu$ is well-defined, and $\mu \otimes \nu = \mu \otimes^l \nu$.*
ii. *Further, if ν is simply a measure on Y, $\mu \otimes^r \nu$ is also well-defined, and $\mu \otimes \nu = \mu \otimes^l \nu = \mu \otimes^r \nu$.*

Characterisations of s-finite kernels v from X to Y ($\neq\emptyset$)

s-finite kernels: T.F.A.E.

1. ν is a s-finite kernel from X to Y
2. $\nu = \sum_{n\in\omega} \nu_n$ for subprobability kernels ν_n.
3. ν is the pushforward of a σ-finite kernel.

   Given kernel $k : X \rightsquigarrow Y$ and measurable function $f : Y \to Z$, define the pushforward kernel $f_* k : X \rightsquigarrow Z$ by: for $x \in X$, $U \in \Sigma_Z$, $f_* k\,(x, U) := k(x, f^{-1}(U))$.

s-finite measures

1. ν is a s-finite measure iff there is a σ-finite measure μ on X and a measurable function $f : X \to \{1, \infty\}$ such that $\nu = \mu(f)$.
2. A weak converse: if ν is an s-finite measure on X, then either ν is zero or $\nu = \mathbb{P}(f)$ for a probability measure $\mathbb{P}$ and a measurable function $f : X \to [0, \infty]$ with $\mathbb{P}([f = 0]) = 0$.

Fix a measure μ on a measurable space (X, Σ_X).

DEF. $U \in \Sigma_X$ is an ∞-set w.r.t. μ if (i) $\mu(U) = \infty$, and (ii) for all $V \in \Sigma_U$, $\mu(V) = 0$ or ∞.

- If μ is σ-finite then it does not have any ∞-sets ($\because$ any set U of infinite μ-measure must have a countable partition of finite μ-measure, i.e., 0).
- IDEA: presence of ∞-sets distinguishes s-finite from σ-finite measures.

Call $U \in \Sigma_X$ a σ-finite complement of X if (i) U is an ∞-set or a null-set, and (ii) μ is σ-finite on $X \setminus U$.

Theorem. Let μ be an s-finite measure on X. Then

i. There exists a σ-finite complement in Σ_X.
ii. μ is σ-finite iff there are no μ-∞-sets in Σ_X.

The converse of (i) fails: not every measure μ which has a σ-finite complement need be s-finite. Take $\mu = \#_{\mathbb{R}} \cdot \infty$.

Outline

Let μ and ν be measures on X. A Radon-Nikodým derivative of μ w.r.t. ν is a measurable function $X \to [0, \infty]$, typically written $\mathrm{d}\mu/\mathrm{d}\nu$, s.t.

$$\mu = \nu\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) := \int_X \nu(\mathrm{d}x)\,\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x). \qquad \text{Notation. } \nu(f) = \int_X \nu(\mathrm{d}x)\, f(x).$$

Recall: μ is absolutely continuous w.r.t. ν (written $\mu \ll \nu$) if $\forall U \in \Sigma_X \,.\, \nu(U) = 0 \implies \mu(U) = 0$.

Theorem (Radon-Nikodým – standard version)

Let $\mu \ll \nu$ be σ-finite measures on a space (X, Σ_X). Then μ has a R-N derivative w.r.t. ν, which is unique up to ν-equivalence.

The pdf of a r.v. is the R-N derivative of the induced measure with respect to some stock measure (usually the Lebesgue measure for continuous r.v.).
- Provides existence proof of conditional expectation for probability measures – key concept in probability theory.
- Basis of compilation of prob. programs to densities (Bhat et al. POPL12; LMCS17; etc.)

Let μ, ν be **s-finite measures** on (X, Σ_X).

DEF. μ is ∞-absolutely continuous w.r.t. ν (written $\mu \overset{\infty}{\ll} \nu$) if (i) $\mu \ll \nu$, and (ii) for all ν-∞-sets U, U is a μ-∞-set or a μ-null-set.

- For σ-finite measure ν, we have $\mu \overset{\infty}{\ll} \nu$ iff $\mu \ll \nu$, vacuously.
- If μ has density f w.r.t. ν (i.e. $\mu = \nu(f)$) then $\mu \overset{\infty}{\ll} \nu$.

DEF. Let $f, g : X \to [0, \infty]$ be measurable, and let $X_\infty \in \Sigma_X$ be a σ-finite complement w.r.t. ν. Say f and g are ν-∞-equivalent if

$$\nu([f \neq g] \cap (X \setminus X_\infty)) + \nu([g = 0 \neq f] \cap X_\infty) + \nu([f = 0 \neq g] \cap X_\infty) \; = \; 0$$

- On the σ-finite part of X: f and g are ν-equivalent
- On σ-finite complement of X: the points where one has value 0 and the other strictly positive are ν-negligible.

Theorem (Radon-Nikodým for s-finite measures)

Let $\mu \overset{\infty}{\ll} \nu$ be s-finite measures on a space (X, Σ_X). Then μ has a R-N derivative w.r.t. ν, which is unique up to ν-∞-equivalence. [False if only assume $\mu \ll \nu$.]

Application

Let $\mu \overset{\infty}{\ll} \nu$ be s-finite measures on X. Then there exists an RN-derivative $\mathrm{d}\mu/\mathrm{d}\nu : X \to [0,\infty]$, satisfying $\nu(\mathrm{d}\mu/\mathrm{d}\nu) = \mu$, unique up to ν-∞-equivalence.

1. Importance sampling of μ w.r.t. ν.

$$\text{sample}\ \mu = \text{let}\ (\text{sample}\ \nu)\ \text{be}\ x\ \text{in}\ \left(\text{score}\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\right); \text{return}\ x\right).$$

2. Rejection sampling of μ w.r.t. ν.
   Assume $\mathrm{d}\mu/\mathrm{d}\nu \leq M \in [0,\infty)$. Let

$$\begin{aligned} f(z) := \text{let}\ (\text{sample}\ \nu)\ \text{be}\ x\ \text{in} \\ (\text{sample}\ \mathbb{U}_{[0,1]})\ \text{be}\ y\ \text{in} \\ \text{if}\ \left(y \leq \frac{1}{M}\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\right)\ \text{then}\ (\text{return}\ x)\ \text{else}\ z. \end{aligned}$$

Then, we get a rejection sampling procedure for μ: sample $\mu = \mathbf{Y}(f)$.

Outline

Disintegration

- Disintegration formalises the idea of a non-trivial "restriction" of a measure to a measure-zero subset of the measure space in question.
- It is closely related to the existence of conditional probability measures.
- Disintegration may be viewed as a process opposite to the construction of a product measure.

  And hence it is related to Fubini theorem.

Borel's Paradox: conditioning on a measure-0 subset

Let x be a point on Earth's surface drawn from a uniform distribution.
If x lies on the equator, its longitude should be uniformly distributed over $[-\pi, \pi]$.

But there is nothing special about the equator: it's just a great circle.
In particular, for a great circle through the poles (i.e. conditioning on the longitude) there should be conditional probability $1/4$ that x lies north of latitude $45^\circ N$.

Now "average out" over the longitude to deduce that x has probability $1/4$ of lying in the spherical cap extending from the north pole down to the 45° parallel of latitude.

Alas, that cap does not cover $1/4$ of the Earth's surface area, as would be required for a point uniformly distributed over Earth's surface.

(Pollard 2002)

Kolmogorov (1930): "The concept of a conditional probability with regard to an isolated hypothesis whose probability equals 0 is inadmissible."

Review: Conditioning for discrete random variables

Straightforward – provided we eschew conditioning on probability-0 events.

DEF. Conditional probability. Assume $\mathbb{P}$ is a probability measure on (X, Σ_X). Suppose r.v. T takes value in $R \subseteq_{\text{fin}} Y$. For $A \in \Sigma_X, y \in R$

$$\mathbb{P}(A \mid T = y) := \frac{\mathbb{P}(A \cap \{T = y\})}{\mathbb{P}(\{T = y\})}$$

Properties: Writing $\mathbb{P}_y(-)$ for the measure $\mathbb{P}(- \mid T = y)$

i. Pre-regularity. $\mathbb{P}_y$ is a probability measure on X, for all $y \in R$.

ii. Concentration. $\mathbb{P}_y$ concentrates on $\{T = y\}$:

$$\mathbb{P}_y(\{T \neq y\}) = \frac{\mathbb{P}(\{T \neq y\} \cap \{T = y\})}{\mathbb{P}(\{T = y\})} = 0.$$

iii. Weighted average. For all $A \in \Sigma_X$, $\mathbb{P}(A) = \sum_{y \in R} \mathbb{P}(\{T = y\})\, \mathbb{P}_y(A)$

Question: How to extend conditional probability $\mathbb{P}(A \mid T = y)$ to general spaces (X, Σ_X) and arbitrary measurable T?

Standard "abstract" approach to conditional expectation

Fundamental Theorem & Definition (Kolmogorov, 1933)

Given triple $(\Omega, \mathcal{F}, \mathbb{P})$, r.v. X with $E(|X|) < \infty$, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. There exists r.v. Y s.t. (i) Y is $\mathcal{G}$ measurable, (ii) $E(|Y|) < \infty$, and (iii) $\forall G \in \mathcal{G} \,.\, \int_G \mathbb{P}(\mathrm{d}\omega)\, Y(\omega) = \int_G \mathbb{P}(\mathrm{d}\omega)\, X(\omega)$.
Moreoever, if Y' is another r.v. satisfying the above, then $Y = Y'$ a.s., and is called a version of conditional expectation $E(X \mid \mathcal{G})$ of X given $\mathcal{G}$.

- There is a gap between intuition and rigour in conditioning arguments.

- An accounting problem: for $F \in \mathcal{F}$ define $\mathbb{P}(F \mid \mathcal{G})$ to be $E(\chi_F \mid \mathcal{G})$.
For a fixed seq. (F_n) of disjoint elts of $\mathcal{F}$, $\mathbb{P}(\bigcup F_n \mid \mathcal{G}) = \sum \mathbb{P}(F_n \mid \mathcal{G})$ a.s.
In general, there are uncountably many such sequences; we cannot conclude ($\because$ uncountably many null-sets) that there is a kernel $P : \Omega \times \mathcal{F} \to [0, 1)$ s.t. (a) $\forall F \in \mathcal{F}$, $P(-, F)$ is (version of) $\mathbb{P}(F \mid \mathcal{G})$, (b) for almost every w, $\mathbb{P}(w, -)$ is a probability measure on $\mathcal{F}$.

Definition: disintegration of a measure

Let $T : X \to Y$ be measurable; μ and ν be measures on X and Y resp.

DEF. A (T, ν)-disintegration (or -conditional distribution) of μ are a family $\{\mu_y\}_{y \in Y}$ of measures on X and a ν-null set $N \in \Sigma_Y$ s.t.

i. Regularity: $(y, U) \mapsto \mu_y(U)$ is a kernel from Y to X;

ii. Concentration: $\forall y \in Y \setminus N$, μ_y concentrates on $\{T = y\}$, i.e., μ_y is supported in $T^{-1}(y)$: $\forall V \in \Sigma_X$, $\mu_y(V) = \mu_y(V \cap T^{-1}(y))$;

iii. Weighted average: $\forall V \in \Sigma_X$, $\mu(V) = \int_Y \nu(\mathrm{d}y)\, \mu_y(V)$.

Often write $\mu(- \mid T = y)$ for μ_y.

iii'. For all measurable $f : X \to [0, \infty]$,

$$\int_X \mu(\mathrm{d}x)\, f(x) = \int_Y \nu(\mathrm{d}y) \int_{T^{-1}(y)} \mu_y(\mathrm{d}x)\, f(x).$$

The standard Disintegration Theorem for σ-finite measures satisfies a weaker definition: if CH holds, $\mu_-(-)$ cannot be a kernel (Maharam's 1950 Problem: Back et al. 2015).

A disintegration theorem for s-finite measures

Let $T : X \to Y$ be measurable from a standard Borel space X to a measurable space Y, let μ and ν resp. be measures on X and Y.

Existence

Assume (i) μ, ν s-finite, (ii) $T_*\mu \ll \nu$, and (iii) for all ν-∞-sets U, $T^{-1}(U)$ is a μ-∞-set or a μ-null-set[a]. Then there exists a (T, ν)-disintegration of μ, $\{\mu_y\}_{y \in Y}$, which is an s-finite kernel.

N.B. Theorem fails for s-finite μ, ν if we only demand $T_*\mu \overset{\infty}{\ll} \nu$.

[a] (ii) and (iii) are strictly stronger than $T_*\mu \overset{\infty}{\ll} \nu$.

Uniqueness

If ν is s-finite, then the (T, ν)-disintegration of μ (*qua* s-finite kernel) is unique up to ν-∞-equivalence.

Fubini's theorem for s-finite measures

Let $\mu = \alpha \otimes \beta$ be a product of s-finite measures on product space $X \times Y$. Let $T : X \times Y \to Y$ be $(x, y) \mapsto y$.

Then the (T, β)-disintegration of μ is $\{\mu_y\}_{y \in Y}$, where $\mu_y = (R_y)_*(\alpha)$ with $R_y : x \mapsto (x, y)$. So μ_y is just a copy of α.

Take measurable $f : X \times Y \to [0, \infty]$. By property (iii) of disintegration:

$$\begin{aligned}
\mu(f) &= \int_Y \beta(\mathrm{d}y) \int_{T^{-1}(y)} \mu_y(\mathrm{d}(x, y))\, f(x, y) \\
&= \int_Y \beta(\mathrm{d}y) \int_{T^{-1}(y)} (R_y)_*(\alpha)(\mathrm{d}(x, y))\, f(x, y) \quad (\because \mu_y = (R_y)_*(\alpha)) \\
&= \int_Y \beta(\mathrm{d}y) \int_X \alpha(\mathrm{d}x)\, \underbrace{f \circ R_y(x)}_{f(x,y)} \quad \text{(by change of variable)}
\end{aligned}$$

which is precisely Fubini's theorem for s-finite measures.

Bayes' Law: Posterior $\propto$ Likelihood $\times$ Prior

$$p(\Theta = \theta \mid X = x) \; = \; \frac{p(X = x \mid \Theta = \theta)\, p(\Theta = \theta)}{p(X = x)}$$

- Bayes' Law says that the posterior times the probability of an observation equals a joint probability.
- But the observation of a continuous quantity usually has probability 0; in which case, Bayes' Law says: "unknown $\times$ 0 = 0"!

(Shan & Ramsey POPL 2017) introduces a new inference algorithm by symbolic manipulation of the prior and an observable expression:

- It can draw exact inference from the observation of a probability-0 continuous quantity.
- Idea: the observable expression denotes a conditional distribution *qua* disintegration of a measure.
- These disintegrations (of s-finite measures) are s-finite kernels, which are denotable by PPL terms.

Problem

Conjecture. Let ρ be an s-finite measure on $X \times Y$ and μ be s-finite measure on X, satisfying condition (C). Then there exists an s-finite kernel $k : X \rightsquigarrow Y$ such that $\rho = \mu \otimes k$. Further the kernel is unique up to μ-∞-equivalence.

Desiderata:

1. Higher order & definability. Take $\mathcal{L}$ an idealised higher-order PPL; e.g. core Hakaru$^{\rightarrow}$(?). Extend ρ and μ to $\mathcal{L}$-definable measures; prove that k is $\mathcal{L}$-definable (Staton ESOP17).
2. Constructiveness / relativised computability. Design an algorithm for constructing k as an $\mathcal{L}$-term, given representations of ρ and μ as $\mathcal{L}$-terms, via partial evaluation (type-directed / continuation-based); prove correctness via synthetic measure theory.
3. Compositionality / "parametricity law". Replace ρ and μ by s-finite kernels (appropriately typed).

Outline

Conclusions and further directions

1. S-finite kernels have good closure properties.
2. Radon-Nikodým and Disintegration theorems extend to s-finite measures.

Further directions

- Methods to construct Radon-Nikodým derivatives and disintegrating measures / kernels
- Deriving disintegration by program transformation & synthesis – an approach to Bayesian inference (Shan & Ramsey, POPL 2017)