New insights into probability on function types

PIHOC 2020, IRIF, Paris

Dario Stein[1], Sam Staton[1], Michael Wolman[2]

February 26, 2020

[1]University of Oxford [2]McGill University

New insights into probability on function types

Aim: Study the nature of probability on function spaces

Outline:

1. The Model
   - Quasi-Borel spaces
   - Descriptive Set Theory
2. A surprising connection
   - H/o probability $\leftrightarrow$ Name generation
3. Structural consequences
   - Non-positive probability

Higher-order probability

General-purpose probabilistic programming:

- Continuous probability distributions $\Rightarrow$ **Meas**urable spaces
- Higher-order constructs are useful
- Compositional semantics? **Meas** is not cartesian closed

Theorem [Aumann'61]

Let $2^{\mathbb{R}}$ denote the space of Borel measurable maps $\mathbb{R} \to 2$. Then there is no σ-algebra on $2^{\mathbb{R}}$ that makes the evaluation map

$$(\ni) : 2^{\mathbb{R}} \times \mathbb{R} \to 2$$

measurable.

Higher-order probability

Some models of higher-order probability

- Spaces of continuous functions
- Measurable cones [Ehrhard,Pagani,Tasson'17]
- Ordered Banach Spaces [Dahlqvist,Kozen'19]
- **Quasi-Borel spaces** [Heunen,Kammar,Staton,Yang'17]

What’s a quasi-Borel space?

Standard Borel spaces

Standard Borel spaces (**Sbs**):

- Well-behaved subcategory of **Meas**

$$S ::= 0 \mid 1 \mid \mathbb{R} \mid \Pi_\omega S \mid \Sigma_\omega S \mid G(S)$$

- Every sbs is countable&discrete or isomorphic to $\mathbb{R}$.

Quasi-Borel spaces

Quasi-Borel spaces (Qbs)

- conservative extension of **Sbs**
- achieve cartesian closure
- nice properties (Fubini, randomization lemma, de Finetti)
- "Denotational Validation of Higher-Order Bayesian Inference" [Ścibior & al.'18]
- "Trace types and denotational semantics for sound programmable inference in probabilistic languages" [Lew & al.'19]

Quasi-Borel spaces

Definition: A qbs is a pair (X, M_X) where $M_X \subseteq [\mathbb{R} \to X]$ is a collection of distinguished maps (satisfying some conditions)

- call $\alpha \in M_X$ "random element"

A morphism $f\colon (X, M_X) \to (Y, M_Y)$ is a map

$$\begin{tikzcd} \mathbb{R} \arrow[d, "\forall \alpha \in M_X"'] \arrow[dr, dotted, "f \circ \alpha \in M_Y"] & \\ X \arrow[r, "f"'] & Y \end{tikzcd}$$

E.g. $M_{\underline{\mathbb{R}}} = \mathbf{Meas}(\mathbb{R}, \mathbb{R})$. Note that $M_X = \mathbf{Qbs}(\underline{\mathbb{R}}, X)$.

Quasi-Borel spaces

There is an idempotent adjunction

$$\begin{tikzcd} \mathbf{Qbs} \arrow[rr, bend left, "\Sigma"] \arrow[rr, phantom, "\perp"] & & \mathbf{Meas} \arrow[ll, bend left, "M"] \end{tikzcd}$$

Where

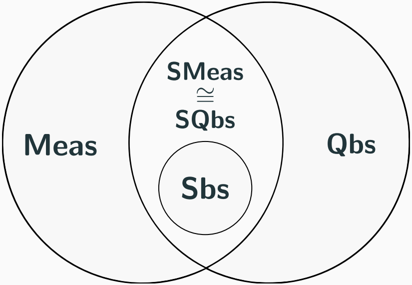
$$M(\Omega) = (|\Omega|, M_\Omega) \qquad M_\Omega = \mathbf{Meas}(\mathbb{R}, \Omega)$$

$$\Sigma(X) = (|X|, \Sigma_X) \qquad \Sigma_X \cong \mathbf{Qbs}(X, 2)$$

$$\Sigma M \Sigma X = \Sigma X \qquad M \Sigma M \Omega = M\Omega$$

Quasi-Borel spaces

We say a qbs is **standard** if its qbs structure comes from a/can be recovered from its σ-algebra.



Qbs conservatively extends **Sbs**

Thm: Function spaces $2^{\mathbb{R}}, \mathbb{R}^{\mathbb{R}}, \ldots$ are *non-standard* qbs

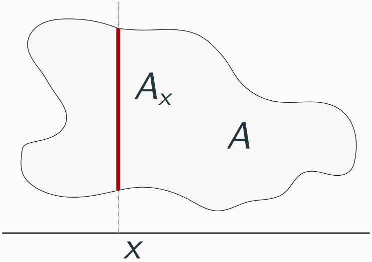
Function spaces

Examples:

- We identify $2^{\mathbb{R}} \cong \mathcal{B}$, the qbs of Borel sets,
- A random element $\mathbb{R} \to 2^{\mathbb{R}}$ must come from currying $A : \mathbb{R} \times \mathbb{R} \to 2$, i.e.

$$x \mapsto A_x = \{y : (x, y) \in A\}$$

for some $A \subseteq \mathbb{R}^2$ Borel.



Function spaces

- Evaluation $(\ni) : 2^{\mathbb{R}} \times \mathbb{R} \to 2$ is a valid morphism
  $\Rightarrow \; (\ni) \in \Sigma_{2^{\mathbb{R}} \times \mathbb{R}}$
- but $(\ni) \notin \Sigma_{2^{\mathbb{R}}} \otimes \Sigma_{\mathbb{R}}$ [Aumann]
  $\Rightarrow \; \Sigma : \mathbf{Qbs} \to \mathbf{Meas}$ does not preserve products

When do we need Σ_X at all?

Measures on qbs

Given a random element $\alpha : \mathbb{R} \to X$, we can pushforward probability from $\mathbb{R}$ to X.

$$P(X) = \{\alpha_* \mu : \alpha \in M_X, \mu \in G(\mathbb{R})\} \subseteq G(X, \Sigma_X).$$

Equality of measures is extensional equality on Σ_X.

- $P(\mathbb{R}) = M(G(\mathbb{R}))$
- P is a strong, affine, commutative monad on **Qbs**

What are distributions on function spaces?

Distributions on function spaces

Easy to use

$$\text{let } a \leftarrow \mathcal{N}(0,1) \text{ in}$$
$$\text{let } b \leftarrow \mathcal{N}(0,1) \text{ in}$$
$$\text{let } f = \lambda x.\, a \cdot x + b \text{ in } \ldots$$
$$\text{observe } y_i \text{ from } \mathcal{N}(f(x_i), \epsilon)$$

Distributions on function spaces

Easy to use

$$\begin{aligned}
&\text{let } a \leftarrow \mathcal{N}(0,1) \text{ in} \\
&\text{let } b \leftarrow \mathcal{N}(0,1) \text{ in} \\
&\text{let } f = \lambda x.\, a \cdot x + b \text{ in } \dots \\
&\text{observe } y_i \text{ from } \mathcal{N}(f(x_i), \epsilon)
\end{aligned}$$

but difficult to analyse directly. So let's do that now!

Crucial example

Random singleton equals emptyset

Theorem (Privacy equation)

Consider the random singleton set

$$X \sim \mathcal{U}[0,1]$$
$$A = \{X\}$$

Random singleton equals emptyset

Theorem (Privacy equation)

Consider the random singleton set

$$X \sim \mathcal{U}[0,1]$$
$$A = \{X\}$$

Then $A \equiv^d \emptyset$.

More formally in $P(2^{\mathbb{R}})$

$$(\text{let } x \leftarrow \mathcal{U}[0,1] \text{ in } \delta(\lambda y.(y = x))) = \delta(\lambda y.\mathrm{false})$$

Privacy equation

Computer scientist (works with name generation): *Not surprised*

Privacy equation [Stark'93]

$$[\![\mathrm{let}\ x = \mathbf{new}\ \mathrm{in}\ \lambda y.(x = y)]\!] = [\![\lambda y.\mathrm{false}]\!]$$

- the name x is *private*
- doesn't get *leaked* from the closure $\lambda y.(x = y)$

But names aren't random numbers, are they?

Privacy equation

Theorem (Privacy equation)

Consider the random singleton set

$$X \sim \mathcal{U}[0,1]$$
$$A = \{X\}$$

Then $A \equiv^d \emptyset$.

Mathematican (surprised) *Wait . . . Surely, every sample of A is non-empty. Can't I tell?*

Privacy equation

Theorem (Privacy equation)

Consider the random singleton set

$$X \sim \mathcal{U}[0, 1]$$
$$A = \{X\}$$

Then $A \equiv^d \emptyset$.

Mathematican (surprised) *Wait ... Surely, every sample of A is non-empty. Can't I tell?*

But can you tell **measurably**?

Measurable properties of functions

What are *measurable properties* of Borel sets?

- morphisms $2^{\mathbb{R}} \to 2$ (second-order type!)

Measurable properties of functions

What are *measurable properties* of Borel sets?

- morphisms $2^{\mathbb{R}} \to 2$ (second-order type!)

Examples: Let $X \sim \mathcal{U}[0,1]$ and $A = \{X\}$

1. membership tests; for any $x_0 \in \mathbb{R}$,

$$x_0 \in A \Leftrightarrow x_0 \in \emptyset \quad \text{a.s.}$$

2. ρ σ-finite measure, then

$$\rho(A) = 0 = \rho(\emptyset) \quad \text{a.s.}$$

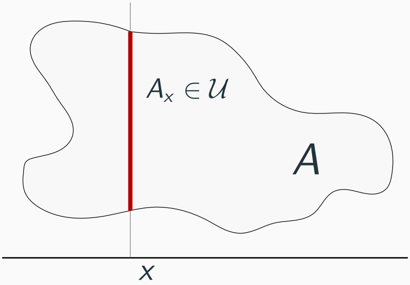
3. But what about checking nonemptyness?

Measurable properties of functions

Borel on Borel [Kechris '87]

Every morphism $\mathcal{U} : 2^{\mathbb{R}} \to 2$ must satisfy

$$\forall A \in \Sigma_{\mathbb{R}^2}, \{x : A_x \in \mathcal{U}\} \text{ Borel.}$$



Measurable properties of functions

Borel on Borel [Kechris '87]

$\mathcal{U}$ Borel on Borel iff $\forall A \in \Sigma_{\mathbb{R}^2}, \{x : A_x \in \mathcal{U}\} \in \Sigma_{\mathbb{R}}$.

Can $\exists : 2^{\mathbb{R}} \to 2$ be a morphism? Then for all $A \subseteq \mathbb{R}^2$ Borel,

$$\pi(A) = \{x : A_x \neq \emptyset\} \text{ must be Borel.}$$

Measurable properties of functions

Borel on Borel [Kechris '87]

$\mathcal{U}$ Borel on Borel iff $\forall A \in \Sigma_{\mathbb{R}^2}, \{x : A_x \in \mathcal{U}\} \in \Sigma_{\mathbb{R}}$.

Can $\exists : 2^{\mathbb{R}} \to 2$ be a morphism? Then for all $A \subseteq \mathbb{R}^2$ Borel,

$$\pi(A) = \{x : A_x \neq \emptyset\} \text{ must be Borel.}$$

Theorem [Lebesgue]: For all $A \subseteq \mathbb{R}^2$ Borel, $\pi(A)$ is Borel.

Measurable properties of functions

Borel on Borel [Kechris '87]

$\mathcal{U}$ Borel on Borel iff $\forall A \in \Sigma_{\mathbb{R}^2}, \{x : A_x \in \mathcal{U}\} \in \Sigma_{\mathbb{R}}$.

Can $\exists : 2^{\mathbb{R}} \to 2$ be a morphism? Then for all $A \subseteq \mathbb{R}^2$ Borel,

$$\pi(A) = \{x : A_x \neq \emptyset\} \text{ must be Borel.}$$

~~**Theorem [Lebesgue]:** For all $A \subseteq \mathbb{R}^2$ Borel, $\pi(A)$ is Borel.~~
Theorem [Suslin]: For $A \subseteq \mathbb{R}^2$ Borel, $\pi(A)$ need not be Borel $\lightning$ (Birthplace of Descriptive Set Theory)

No $\mathcal{U} : 2^{\mathbb{R}} \to 2$ can distinguish between $\{X\}$ and $\emptyset$ with positive probability

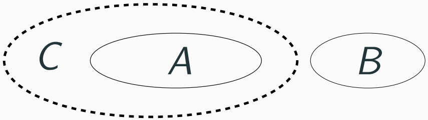
Into descriptive set theory

Theorem

For all Borel on Borel $\mathcal{U}$, $\emptyset \in \mathcal{U} \Leftrightarrow \{x\} \in \mathcal{U}$ for almost all x.

Idea "Borel inseparability".

- A, B are *Borel inseparable* if there is no Borel C with



- There is a Borel set $C \subseteq \mathbb{R}^2$ such that $C^0 = \{x : C_x \text{ empty}\}$ and $C^1 = \{x : C_x \text{ singleton}\}$ are Borel inseparable [Becker].

Into descriptive set theory

Theorem

For all Borel on Borel $\mathcal{U}$, $\emptyset \in \mathcal{U} \Leftrightarrow \{x\} \in \mathcal{U}$ for almost all x.

Sketch. Assume $\emptyset \in \mathcal{U}$ but $S = \{x : \{x\} \notin \mathcal{U}\}$ has positive measure. Do some encoding to let Becker's set C lie in $\mathbb{R} \times S$. Then $B = \{x : C_x \in \mathcal{U}\}$ is Borel and

1. if $x \in C^0$ then $C_x = \emptyset \in \mathcal{U}$, so $x \in B$.
2. if $x \in C^1$ then $C_x = \{s\}$ for some $s \in S$, so $x \notin B$.

Thus B separates C^0 and C^1 $\lightning$

Generalizing

Random transposition = identity

Consider the transposition map $\tau : \mathbb{R}^2 \to \mathbb{R}^{\mathbb{R}}$

$$\tau(a, b)(x) = (a\, b)(x).$$

Then we have

$$\big(\text{let } (a, b) \leftarrow \mathcal{U}[0, 1]^2 \text{ in } \delta(\tau(a, b))\big) = \delta(\mathrm{id}_{\mathbb{R}}) \in P(\mathbb{R}^{\mathbb{R}})$$

Descriptive Set Theory: More sophisticated encoding
Name-generation: Swapping two private names is not observable

Names & Probability

Name generation & probability

Name-generation is a synthetic* probabilistic effect.

- commutative & discardable
- models: e.g nominal sets & name-generation monad [Stark'96, Pitts'13]

We can interpret it as an actual probabilistic effect.

Theorem

Higher-order PPLs are a sound and correct models for Stark's ν-calculus

1. names are interpreted in $\mathbb{R}$
2. name-generation is sampling a continuous distribution

Stark's ν-calculus

Name ideas inevitably show up in higher-order PPL, but

Name generation is subtle

$$\nu x.\lambda y.x \quad \not\approx \quad \lambda y.\nu x.x$$

$$\nu a.\nu b.\lambda x.\text{if } (x = a) \text{ then } a \text{ else } b \quad \approx \quad \nu b.\lambda x.b$$

$$\nu a.\nu b.\lambda x.\text{if } (x = b) \text{ then } a \text{ else } b \quad \not\approx \quad \nu b.\lambda x.b$$

Which equivalences are verified in probabilistic semantics?

Full abstraction problem

Theorem

Let M, M' be ν-calculus expressions of type

- $\tau_1 \to \cdots \to \tau_n \to \mathrm{bool}$
- or $\tau_1, \ldots, \tau_n \to \mathrm{name}$, $\tau_i \in \{\mathrm{bool}, \mathrm{name}\}$

then $M \approx M' \Leftrightarrow [\![M]\!] = [\![M']\!]$ in **Qbs**.

Conjecture

Full abstraction at all iterated function types

$$\tau_1 \to \cdots \tau_n \to \tau$$

This is **more abstract** than traditional semantics! (nominal sets don't validate the Privacy equation).

Structural consequences

Structural consequences

Synthetic probability theory [Fritz'19]

- categorical axiomatization of probabilistic systems
- high-level comparison of properties

Structural consequences

Synthetic probability theory [Fritz'19]

- categorical axiomatization of probabilistic systems
- high-level comparison of properties

Example: Deterministic marginals

Given a joint distribution (X, Y) with X deterministic. Then X and Y are independent.

- True for discrete probability
- True on **Meas** (product-σ-algebras!)
- blantantly fails with negative probabilities ($D_\pm$)
- axiomatized by a property called "positivity"

Non-positivity

Name generation is non-positive

Name-generation violates deterministic marginals:

$$[\![\text{let } x = \textbf{new} \text{ in } (\lambda y.(y = x), x)]\!] \in T(2^{\mathbb{A}} \times \mathbb{A})$$

By Privacy equation:

- first marginal is deterministically $\lambda y.\mathrm{false}$.
- **not independent** of x, which is *leaked*

Qbs is non-positive for the same reason

- requires failure of product-preservation (**Meas** is positive)
- this shows $\Sigma(2^{\mathbb{R}} \times \mathbb{R}) \neq \Sigma(2^{\mathbb{R}}) \otimes \Sigma(\mathbb{R})$ [Aumann]

Conclusion

1. **Qbs** is a convenient category to work in
    - Usual probability theory at ground types
    - Descriptive set theory at function types
    - Random singleton $= \emptyset$
    - **Conjecture:** Full abstraction at first-order for ν-calculus
      (Already more abstract than nominal sets)

Conclusion

1. **Qbs** is a convenient category to work in
   - Usual probability theory at ground types
   - Descriptive set theory at function types
   - Random singleton $= \emptyset$
   - **Conjecture:** Full abstraction at first-order for ν-calculus (Already more abstract than nominal sets)
2. **Higher-order probability** (model independent)
   - Measures on function types are interesting to study
   - Inevitable connection with name generation
   - H/o measurability $\leftrightarrow$ second-order programs $2^{\mathbb{R}} \to 2$
   - Non-positivity is a feature
   - Randomization is anonymization (diff. privacy)

Takeaway

If you have a model of higher-order probability supporting

1. continuous distributions
2. equality checks $\mathbb{R} \times \mathbb{R} \to 2$

$\Rightarrow$ **Test it against ν-calculus and tell me what happens!**