Towards Minimal Axioms for Probabilistic Programming

Eigil Fjeldgren Rischel

MSFP, 2026

Tallinn University of Technology

Introduction

Probabilistic programming: programs that do random things.

Introduction

Probabilistic programming: programs that do random things.

You model this by putting a monad P on the category of contexts/terms, where a value of type $P(X)$ is a random value of type X. (In other words: treat stochasticity as an effect). But which monad?

Introduction

Probabilistic programming: programs that do random things.

You model this by putting a monad P on the category of contexts/terms, where a value of type $P(X)$ is a random value of type X. (In other words: treat stochasticity as an effect). But which monad?

What does this monad need to be a good model of probability?

Introduction

Related question: what is special about the *Giry monad* $G : \mathrm{Meas} \to \mathrm{Meas}$ on measurable spaces, where GX is the space of bona fide probability measures on X?

Introduction

Related question: what is special about the *Giry monad* $G : \mathrm{Meas} \to \mathrm{Meas}$ on measurable spaces, where GX is the space of bona fide probability measures on X?

Interesting on its own, may also give us a good idea about what to do for programming.

Introduction

Related question: what is special about the *Giry monad* $G : \mathbf{Meas} \to \mathbf{Meas}$ on measurable spaces, where GX is the space of bona fide probability measures on X?

Interesting on its own, may also give us a good idea about what to do for programming.

We restrict to the class of *standard Borel* measurable spaces and consider instead the restricted $G : \mathbf{Borel} \to \mathbf{Borel}$.

Main theorem

Theorem

$G : \mathrm{Borel} \to \mathrm{Borel}$ *is initial among affine commutative monads which preserve countable cofiltered limits, pullbacks along coproduct inclusions, and contain a commutative binary operation* $\mathsf{cf} : G(2)$.

Main theorem

Theorem

$G : \mathrm{Borel} \to \mathrm{Borel}$ *is initial among affine commutative monads which preserve countable cofiltered limits, pullbacks along coproduct inclusions, and contain a commutative binary operation* $\mathsf{cf} : G(2)$.

This is essentially a *completeness result*—we can construct every distribution, and prove every identity between distributions, using these axioms.

Main theorem

Theorem

$G : \mathrm{Borel} \to \mathrm{Borel}$ is initial among affine commutative monads which preserve countable cofiltered limits, pullbacks along coproduct inclusions, and contain a commutative binary operation $\mathbf{cf} : G(2)$.

This is essentially a *completeness result*—we can construct every distribution, and prove every identity between distributions, using these axioms.

Let's consider these properties in turn. First, **cf** is just the uniform distribution on two points.

Pullbacks

G preserves pullbacks along coproduct inclusions: Given $f : X \to A + B$, a distribution on $f^{-1}(A)$ is equivalently a distribution on X so that the image is concentrated on A.

$$\begin{array}{ccc} G(f^{-1}(A)) & \longrightarrow & G(X) \\ \downarrow & & \downarrow f \\ G(A) & \longrightarrow & G(A + B) \end{array}$$

Cofiltered limits

G preserves cofiltered limits: a distribution on sequences, $p \in G(A^{\mathbb{N}})$, is equivalent to a compatible family of finite-dimensional distributions $p_n \in G(A^n)$ for $n \in \mathbb{N}$.

Cofiltered limits

G preserves cofiltered limits: a distribution on sequences, $p \in G(A^{\mathbb{N}})$, is equivalent to a compatible family of finite-dimensional distributions $p_n \in G(A^n)$ for $n \in \mathbb{N}$.This is *Kolmogorov's extension theorem*. This is true for standard Borel spaces but not for general measurable spaces.

Proof outline

Let P be a monad satisfying the hypotheses of the theorem.

Proof outline

Let P be a monad satisfying the hypotheses of the theorem.

- By the extension property, obtain an infinite random bitstream $\mathsf{cf}^{\mathbb{N}} : P(2^{\mathbb{N}})$.

Proof outline

Let P be a monad satisfying the hypotheses of the theorem.

- By the extension property, obtain an infinite random bitstream $\mathbf{cf}^{\mathbb{N}} : P(2^{\mathbb{N}})$.
- Represent every countably supported distribution in terms of this.

Proof outline

Let P be a monad satisfying the hypotheses of the theorem.

- By the extension property, obtain an infinite random bitstream $\mathbf{cf}^{\mathbb{N}} : P(2^{\mathbb{N}})$.
- Represent every countably supported distribution in terms of this.
- Hard part: Prove that this can be done in a homomorphic way, yielding a monad homomorphism $\Delta \to P$.

Proof outline

Let P be a monad satisfying the hypotheses of the theorem.

- By the extension property, obtain an infinite random bitstream $\mathbf{cf}^{\mathbb{N}} : P(2^{\mathbb{N}})$.
- Represent every countably supported distribution in terms of this.
- Hard part: Prove that this can be done in a homomorphic way, yielding a monad homomorphism $\Delta \to P$.
- Kuratowski's theorem: Every standard Borel space is countable or isomorphic to $2^{\mathbb{N}}$. Hence by the extension property, there's an most one way to extend this along $\Delta \hookrightarrow G$. Check that this is indeed a monad homomorphism.

Proof outline, cont.

The hard part is to prove identities between different constructions of the same measure in terms of infinite coinflip sequences.

Proof outline, cont.

The hard part is to prove identities between different constructions of the same measure in terms of infinite coinflip sequences.

The actual proof is very fiddly. The main lever is this:

Lemma

Let $V_1 \subseteq 2^{\mathbb{N}}$ *be the subspace of sequences that contain infinitely many* 1*s. Then* $\mathbf{cf}^{\mathbb{N}}$ *lifts to* $P(V_1)$*.*

Proof outline, cont.

The hard part is to prove identities between different constructions of the same measure in terms of infinite coinflip sequences.

The actual proof is very fiddly. The main lever is this:

Lemma

Let $V_1 \subseteq 2^{\mathbb{N}}$ *be the subspace of sequences that contain infinitely many* 1*s. Then* $\mathbf{cf}^{\mathbb{N}}$ *lifts to* $P(V_1)$*.*

To see this for actual probability distributions, you would normally use a quantitative argument (the probability of no ones is $< 1/2^n$ for every n...).

Proof outline, cont.

The hard part is to prove identities between different constructions of the same measure in terms of infinite coinflip sequences.

The actual proof is very fiddly. The main lever is this:

Lemma

Let $V_1 \subseteq 2^{\mathbb{N}}$ *be the subspace of sequences that contain infinitely many* 1*s. Then* $\mathbf{cf}^{\mathbb{N}}$ *lifts to* $P(V_1)$*.*

To see this for actual probability distributions, you would normally use a quantitative argument (the probability of no ones is $< 1/2^n$ for every n...).Surprisingly, follows from our purely qualitative hypotheses.

Proof of lemma

Let $\chi : 2^{\mathbb{N}} \to 2$ be the indicator of V_1. Note it is independent of every finite prefix.

Proof of lemma

Let $\chi : 2^{\mathbb{N}} \to 2$ be the indicator of V_1. Note it is independent of every finite prefix.This implies $P(\chi)\,\mathrm{cf}^{\mathbb{N}} \in P(2)$ is *deterministic*—that is, its pairing with itself in $P(2 \times 2)$ factors over the diagonal.

Proof of lemma

Let $\chi : 2^{\mathbb{N}} \to 2$ be the indicator of V_1. Note it is independent of every finite prefix.This implies $P(\chi)\,\mathrm{cf}^{\mathbb{N}} \in P(2)$ is *deterministic*—that is, its pairing with itself in $P(2 \times 2)$ factors over the diagonal.

This + symmetry of $\mathbf{cf}$ implies that

$$P(\Delta_2\chi)\,\mathrm{cf}^{\mathbb{N}} = P(\chi \times \chi)(\mathrm{cf}^{\mathbb{N}} \otimes \mathrm{cf}^{\mathbb{N}}) = P(\chi \times (\chi\neg))(\mathrm{cf}^{\mathbb{N}} \otimes \mathrm{cf}^{\mathbb{N}}) = P(\chi \times \chi\neg)P(\Delta_{2^{\mathbb{N}}})\,\mathrm{cf}^{\mathbb{N}}$$

Proof of lemma

Let $\chi : 2^{\mathbb{N}} \to 2$ be the indicator of V_1. Note it is independent of every finite prefix.This implies $P(\chi)\,\mathrm{cf}^{\mathbb{N}} \in P(2)$ is *deterministic*—that is, its pairing with itself in $P(2 \times 2)$ factors over the diagonal.

This + symmetry of $\mathbf{cf}$ implies that

$$P(\Delta_2 \chi)\,\mathrm{cf}^{\mathbb{N}} = P(\chi \times \chi)(\mathrm{cf}^{\mathbb{N}} \otimes \mathrm{cf}^{\mathbb{N}}) = P(\chi \times (\chi\neg))(\mathrm{cf}^{\mathbb{N}} \otimes \mathrm{cf}^{\mathbb{N}}) = P(\chi \times \chi\neg)P(\Delta_{2^{\mathbb{N}}})\,\mathrm{cf}^{\mathbb{N}}$$

By postcomposing $\mathbf{v}$, we show that $P(\chi)\,\mathrm{cf}^{\mathbb{N}} = \eta$. This + the fact that P preserves pullbacks finishes the proof.

What about programming?

Can we program like this?

What about programming?

Can we program like this?

In a dependently typed language, we can internalize the limits involved in this theorem, and add operators implementing the required isomorphism.

$$\mathsf{ext} : \sum_{p : \Pi_{n:\mathbb{N}} P(X^n)} \Pi_{n:\mathbb{N}} P(\pi_{1,\ldots n-1})(pn) =_{P(X^{n-1})} p(n-1) \to P(X^{\mathbb{N}})$$

(We could also replace $X^{\mathbb{N}}$ with a dependent function type here, of course).

What about programming?

Can we program like this?

In a dependently typed language, we can internalize the limits involved in this theorem, and add operators implementing the required isomorphism.

$$\mathsf{ext} : \sum_{p : \Pi_{n:\mathbb{N}} P(X^n)} \Pi_{n:\mathbb{N}} P(\pi_{1,\ldots n-1})(pn) =_{P(X^{n-1})} p(n-1) \to P(X^{\mathbb{N}})$$

(We could also replace $X^{\mathbb{N}}$ with a dependent function type here, of course).

The question is

- Does this allow us to express useful probabilistic programs?
- Can we prove a reasonable number of equations between them?

Theorem

Let $\mathcal{C}$ be an elementary topos with countable choice, and let $P : \mathcal{C} \to \mathcal{C}$ be an affine commutative monad which preserves countable cofiltered limits and pullbacks along all monomorphisms, and which has a commutative binary operation. Then there exists a map $b : [0,1] \to P(2)$ so that the induced $[0,1] \times P(X)^2 \to P(X)$ makes every $P(X)$ into an internal convex space.

(This is work in progress)

Theorem

Let C be an elementary topos with countable choice, and let $P : C \to C$ be an affine commutative monad which preserves countable cofiltered limits and pullbacks along all monomorphisms, and which has a commutative binary operation. Then there exists a map $b : [0,1] \to P(2)$ so that the induced $[0,1] \times P(X)^2 \to P(X)$ makes every $P(X)$ into an internal convex space.

(This is work in progress) In other words, we can define the Bernoulli distribution $b(\lambda) \in P(2)$ for each $\lambda \in [0,1]$, and the expected equations between finitary distributions defined in terms of these hold.

Theorem

Let C be an elementary topos with countable choice, and let $P : C \to C$ be an affine commutative monad which preserves countable cofiltered limits and pullbacks along all monomorphisms, and which has a commutative binary operation. Then there exists a map $b : [0,1] \to P(2)$ so that the induced $[0,1] \times P(X)^2 \to P(X)$ makes every $P(X)$ into an internal convex space.

(This is work in progress) In other words, we can define the Bernoulli distribution $b(\lambda) \in P(2)$ for each $\lambda \in [0,1]$, and the expected equations between finitary distributions defined in terms of these hold. (This may hold without countable choice, with the Cauchy reals.)

Conclusion

In a sufficiently rich ambient language, the single distribution $\mathbf{cf} : P(2)$, along with two hypotheses about preservation of limits, suffices to generate every distribution of interest.

Conclusion

In a sufficiently rich ambient language, the single distribution $\mathbf{cf} : P(2)$, along with two hypotheses about preservation of limits, suffices to generate every distribution of interest.*And to prove* at least a large set of equations between such distributions.

Conclusion

In a sufficiently rich ambient language, the single distribution $\mathsf{cf} : P(2)$, along with two hypotheses about preservation of limits, suffices to generate every distribution of interest.*And to prove* at least a large set of equations between such distributions.

Ongoing questions:

- Under what assumption on C can we prove which identities between distributions using only these assumptions on P?

Conclusion

In a sufficiently rich ambient language, the single distribution $\mathsf{cf} : P(2)$, along with two hypotheses about preservation of limits, suffices to generate every distribution of interest.*And to prove* at least a large set of equations between such distributions.

Ongoing questions:

- Under what assumption on C can we prove which identities between distributions using only these assumptions on P?
- Can we describe the initial monad in a useful way?

Conclusion

In a sufficiently rich ambient language, the single distribution $\mathsf{cf} : P(2)$, along with two hypotheses about preservation of limits, suffices to generate every distribution of interest.*And to prove* at least a large set of equations between such distributions.

Ongoing questions:

- Under what assumption on C can we prove which identities between distributions using only these assumptions on P?
- Can we describe the initial monad in a useful way?
- Is there a useful version of this without dependent types?

Thank you for listening.

Paper: “The Universal Property of Measure-Theoretic Probability”, arxiv:2512.15485