

# Probabilistic Functional Programming

---

Donnacha Oisín Kidney

July 18, 2018

Modeling Probability

An Example

Unclear Semantics

Underpowered

Monadic Modeling

The Erwig And

Kollmansberger Approach

Other Interpreters

Theoretical Foundations

Stochastic Lambda

Calculus

Giry Monad

Other Applications

Differential Privacy

Conclusion

# Modeling Probability

---

How do we model stochastic and probabilistic processes in programming languages?

# The Boy-Girl Paradox<sup>1</sup>

(apologies for the outdated language)

1. Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?
2. Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

---

<sup>1</sup>Martin Gardner. *The 2nd Scientific American Book of Mathematical Puzzles & Diversions*. University of Chicago Press ed. Chicago: University of Chicago Press, 1987. ISBN: 978-0-226-28253-4.

# The Boy-Girl Paradox<sup>1</sup>

(apologies for the outdated language)

1. Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?
2. Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

Is the answer to 2  $\frac{1}{3}$  or  $\frac{1}{2}$ ?

---

<sup>1</sup>Martin Gardner. *The 2nd Scientific American Book of Mathematical Puzzles & Diversions*. University of Chicago Press ed. Chicago: University of Chicago Press, 1987. ISBN: 978-0-226-28253-4.

# The Boy-Girl Paradox<sup>1</sup>

(apologies for the outdated language)

1. Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?
2. Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

Is the answer to 2  $\frac{1}{3}$  or  $\frac{1}{2}$ ?

Part of the difficulty in the question is that it's ambiguous: can we use programming languages to lend some precision?

---

<sup>1</sup>Martin Gardner. *The 2nd Scientific American Book of Mathematical Puzzles & Diversions*. University of Chicago Press ed. Chicago: University of Chicago Press, 1987. ISBN: 978-0-226-28253-4.

Gardner originally wrote that the second question (perhaps surprisingly) has the answer  $\frac{1}{3}$ . However, he later acknowledged the question was ambiguous, and agreed that certain interpretations could correctly conclude its answer was  $\frac{1}{2}$ .



Using normal features built in to the language.

```
from random import randrange, choice

class Child:
    def __init__(self):
        self.gender = choice(['boy', 'girl'])
        self.age = randrange(18)
```

```
from operator import attrgetter

def mr_jones():
    child_1 = Child()
    child_2 = Child()
    eldest = max(child_1, child_2,
                  key=attrgetter('age'))
    assert eldest.gender == 'girl'
    return [child_1, child_2]
```

```
def mr_smith():  
    child_1 = Child()  
    child_2 = Child()  
    assert child_1.gender == 'boy' or \  
           child_2.gender == 'boy'  
    return [child_1, child_2]
```

# Unclear semantics

What contracts are guaranteed by probabilistic functions?

What does it mean *exactly* for a function to be probabilistic?

Why isn't the following<sup>2</sup> “random”?

```
int getRandomNumber()  
{  
    return 4; // chosen by fair dice roll.  
             // guaranteed to be random.  
}
```

---

<sup>2</sup>Randall Munroe. *Xkcd: Random Number*. en. Title text: RFC 1149.5 specifies 4 as the standard IEEE-vetted random number. Feb. 2007. URL: <https://xkcd.com/221/> (visited on 07/06/2018).

What about this?

```
children_1 = [Child(), Child()]  
children_2 = [Child()] * 2
```

How can we describe the difference between `children_1` and `children_2`?

# Probabilistic Functional Programming

- └ Modeling Probability

- └ Unclear Semantics

What about this?

```
children_1 = [Child(), Child()]  
children_2 = [Child()] * 2
```

How can we describe the difference between `children_1` and `children_2`?

The first runs two random processes; the second only one. Both have the same types, both look like they do the same thing. We need a good way to describe the difference between them.

## Underpowered

There are many more things we may want to do with probability distributions.

What about expectations?

```
def expect(predicate, process, iterations=100):
    success, tot = 0, 0
    for _ in range(iterations):
        try:
            success += predicate(process())
            tot += 1
        except AssertionError:
            pass
    return success / tot
```

2018-07-18

# Probabilistic Functional Programming

└ Modeling Probability

└ Underpowered

└ Underpowered

Underpowered

There are many more things we may want to do with probability distributions.

What about expectations?

```
def expect(predicate, process, iterations=100):
    success, tot = 0, 0
    for _ in range(iterations):
        try:
            success += predicate(process())
            tot += 1
        except AssertionError:
            pass
    return success / tot
```

This solution is both inefficient and inexact. Also, we may want to express other attributes of probability distributions: independence, for example.



## The Ad-Hoc Solution

```
p_1 = expect(  
    lambda children: all(child.gender == 'girl'  
                          for child in children),  
    mr_jones)  
p_2 = expect(  
    lambda children: all(child.gender == 'boy'  
                          for child in children),  
    mr_smith)
```

$$p_1 \approx \frac{1}{2}$$

$$p_2 \approx \frac{1}{3}$$

# Monadic Modeling

---

What we're approaching is a DSL, albeit an unspecified one.

What we're approaching is a DSL, albeit an unspecified one.

Three questions for this DSL:

What we're approaching is a DSL, albeit an unspecified one.

Three questions for this DSL:

- Why should we implement it? What is it useful for?

What we're approaching is a DSL, albeit an unspecified one.

Three questions for this DSL:

- Why should we implement it? What is it useful for?
- How should we implement it? How can it be made efficient?

What we're approaching is a DSL, albeit an unspecified one.

Three questions for this DSL:

- Why should we implement it? What is it useful for?
- How should we implement it? How can it be made efficient?
- Can we glean any insights on the nature of probabilistic computations from the language? Are there any interesting symmetries?

# The Erwig And Kollmansberger Approach

First approach<sup>3</sup>:

```
newtype Dist a = Dist { runDist :: [(a, ℝ)] }
```

A distribution is a list of possible events, each tagged with a probability.

---

<sup>3</sup>Martin Erwig and Steve Kollmansberger. “Functional Pearls: Probabilistic Functional Programming in Haskell”. In: *Journal of Functional Programming* 16.1 (2006), pp. 21–34. ISSN: 1469-7653, 0956-7968. DOI: [10.1017/S0956796805005721](https://doi.org/10.1017/S0956796805005721). URL: <http://web.engr.oregonstate.edu/~erwig/papers/abstracts.html%5C#JFP06a> (visited on 09/29/2016).



2018-07-18

# Probabilistic Functional Programming

- └ Monadic Modeling

- └ The Erwig And Kollmansberger Approach

- └ The Erwig And Kollmansberger Approach

First approach<sup>1</sup>:

```
newtype Dist a = Dist {runDist :: [(a, R)]}
```

A distribution is a list of possible events, each tagged with a probability.

<sup>1</sup>Martin Erwig and Steve Kollmansberger. "Functional Pearls: Probabilistic Functional Programming in Haskell". In: *Journal of Functional Programming* 16:1 (2006), pp. 27–34. ISSN: 1469-7603, 0958-7988. DOI: 10.1017/S0958790805005723. URL: <http://web.engr.oregonstate.edu/~erwig/papers/abstracts.html#SC03FP06a> (retrieved on 07/20/2018).

This representation only works for discrete distributions

We could (for example) encode a die as:

*die* :: *Dist Integer*

*die* = *Dist* [(1,  $\frac{1}{6}$ ), (2,  $\frac{1}{6}$ ), (3,  $\frac{1}{6}$ ), (4,  $\frac{1}{6}$ ), (5,  $\frac{1}{6}$ ), (6,  $\frac{1}{6}$ )]

This lets us encode (in the types) the difference between:

*children\_1* :: [*Dist Child*]

*children\_2* :: *Dist* [*Child*]

As we will use this as a DSL, we need to define the language features we used above:

```
def mr_smith():  
    child_1 = Child()  
    child_2 = Child()  
    assert child_1.gender == 'boy' or \  
           child_2.gender == 'boy'  
    return [child_1, child_2]
```

As we will use this as a DSL, we need to define the language features we used above:

```
def mr_smith():  
    child_1 = Child()  
    child_2 = Child()  
    assert child_1.gender == 'boy' or \  
           child_2.gender == 'boy'  
    return [child_1, child_2]
```

1. = (assignment)

As we will use this as a DSL, we need to define the language features we used above:

```
def mr_smith():  
    child_1 = Child()  
    child_2 = Child()  
    assert child_1.gender == 'boy' or \  
           child_2.gender == 'boy'  
    return [child_1, child_2]
```

1. = (assignment)
2. assert

As we will use this as a DSL, we need to define the language features we used above:

```
def mr_smith():  
    child_1 = Child()  
    child_2 = Child()  
    assert child_1.gender == 'boy' or \  
           child_2.gender == 'boy'  
    return [child_1, child_2]
```

1. = (assignment)
2. assert
3. return

## Assignment i

Assignment expressions can be translated into lambda expressions:

$$\begin{aligned} & \text{let } x = e_1 \text{ in } e_2 \\ \equiv & \\ & (\lambda x. e_2) e_1 \end{aligned}$$

In the context of a probabilistic language,  $e_1$  and  $e_2$  are distributions. So what we need to define is application: this is encapsulated by the “monadic bind”:

$$(\gg) :: \text{Dist } a \rightarrow (a \rightarrow \text{Dist } b) \rightarrow \text{Dist } b$$



## Assignment ii

For a distribution, what's happening inside the  $\lambda$  is  $e_1$  given  $x$ . Therefore, the resulting probability is the product of the outer and inner probabilities.

$$xs \gg f = \text{Dist} \left[ \begin{array}{l} (y, xp \times yp) \\ | (x, xp) \leftarrow \text{runDist } xs \\ , (y, yp) \leftarrow \text{runDist } (f \ x) \end{array} \right]$$

# Assertion

Assertion is a kind of conditioning: given a statement about an event, it either occurs or it doesn't.

```
guard :: Bool → Dist ()  
guard True = Dist [((), 1)]  
guard False = Dist []
```

# Return

Return is the “unit” value for a distribution; the certain event, the unconditional distribution.

*return* ::  $a \rightarrow \text{Dist } a$

*return*  $x = \text{Dist } [(x, 1)]$

## Putting it all Together

$mrSmith :: Dist [Child]$

$mrSmith = do$

$child_1 \leftarrow child$

$child_2 \leftarrow child$

$guard (gender\ child_1 \equiv Boy \vee gender\ child_2 \equiv Boy)$

$return [child_1, child_2]$

$expect :: (a \rightarrow \mathbb{R}) \rightarrow Dist\ a \rightarrow \mathbb{R}$

$expect\ p\ xs = \frac{sum [p\ x \times xp | (x, xp) \leftarrow runDist\ xs]}{sum [xp | (-, xp) \leftarrow runDist\ xs]}$

$probOf :: (a \rightarrow Bool) \rightarrow Dist\ a \rightarrow \mathbb{R}$

$probOf\ p = expect\ (\lambda x \rightarrow \text{if } p\ x \text{ then } 1 \text{ else } 0)$

$\text{probOf } (\text{all } ((\equiv) \text{ Girl } \circ \text{ gender})) \text{ mrJones} \equiv \frac{1}{2}$   
 $\text{probOf } (\text{all } ((\equiv) \text{ Boy } \circ \text{ gender})) \text{ mrSmith} \equiv \frac{1}{3}$

# Alternative Interpreters

Once the semantics are described, different interpreters are easy to swap in.

```
data Decision = Decision {stick :: Bool  
                          , switch :: Bool}
```

```
montyHall :: Dist Decision
```

```
montyHall = do
```

```
  car      ← uniform [1..3]
```

```
  choice1 ← uniform [1..3]
```

```
  let left  = [door | door ← [1..3], door ≠ choice1]
```

```
  let open  = head [door | door ← left, door ≠ car]
```

```
  let choice2 = head [door | door ← left, door ≠ open]
```

```
  return (Decision {stick  = car ≡ choice1  
                  , switch = car ≡ choice2})
```

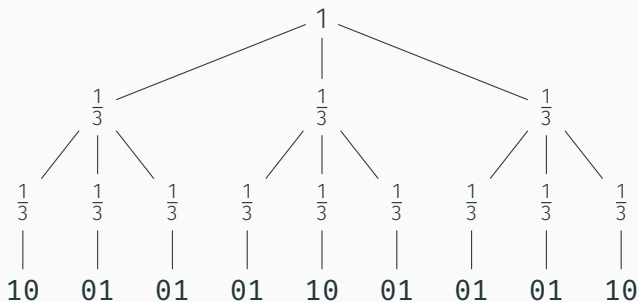
While we can interpret it in the normal way to solve the problem:

$$\begin{aligned} \text{probOf stick montyHall} &\equiv \frac{1}{3} \\ \text{probOf switch montyHall} &\equiv \frac{2}{3} \end{aligned}$$



## Monty Hall iii

We could alternatively draw a diagram of the process.



**Figure 1:** AST from Monty Hall problem. **1** is a win, **0** is a loss. The first column is what happens on a stick, the second is what happens on a loss.

# Theoretical Foundations

---

It is possible<sup>4</sup> to give measure-theoretic meanings to the operations described above.

$$\mathcal{M} \llbracket \text{return } x \rrbracket (A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$\mathcal{M} \llbracket d \ggg k \rrbracket (A) = \int_X \mathcal{M} \llbracket k(x) \rrbracket (A) d\mathcal{M} \llbracket d \rrbracket (x) \quad (2)$$

---

<sup>4</sup>Norman Ramsey and Avi Pfeffer. “Stochastic Lambda Calculus and Monads of Probability Distributions”. In: *29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. Vol. 37. ACM, 2002, pp. 154–165. URL: <http://www.cs.tufts.edu/~nr/cs257/archive/norman-ramsey/pmonad.pdf> (visited on 09/29/2016).

2018-07-18

# Probabilistic Functional Programming

- └ Theoretical Foundations

- └ Stochastic Lambda Calculus

- └ Stochastic Lambda Calculus

It is possible<sup>8</sup> to give measure-theoretic meanings to the operations described above.

$$\mathcal{M}[\text{return } x](A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$\mathcal{M}[d \gg= h](A) = \int_X \mathcal{M}[h(x)](A) d\mathcal{M}[d](x) \quad (2)$$

<sup>8</sup>Norman Ramsey and An Pfaffel, "Stochastic Lambda Calculus and Moments of Probability Distributions", in 29th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, Vol. 31 ACM, 2002, pp. 704-705. URL: <http://www.cs.tufts.edu/~npr/cs237/archive/norman-ramsey/pmsad.pdf> (visited on 05/26/2016).

return is the Dirac measure

Giry<sup>5</sup> gave a categorical interpretation of probability theory.

---

<sup>5</sup>Michèle Giry. “A Categorical Approach to Probability Theory”. In: *Categorical Aspects of Topology and Analysis*. Ed. by A. Dold, B. Eckmann, and B. Banaschewski. Vol. 915. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 68–85. ISBN: 978-3-540-11211-2 978-3-540-39041-1. DOI: [10.1007/BFb0092872](https://doi.org/10.1007/BFb0092872). URL: <http://link.springer.com/10.1007/BFb0092872> (visited on 03/03/2017).

## Categories, Quickly

## Categories, Quickly

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow^{g \circ f} & \downarrow g \\ & & Z \end{array}$$

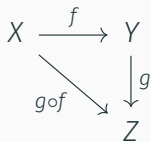
## Categories, Quickly

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow^{g \circ f} & \downarrow g \\ & & Z \end{array}$$

Objects  $\text{Ob}(\mathcal{C}) = \{X, Y, Z\}$



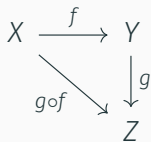
## Categories, Quickly



Objects  $\text{Ob}(\mathcal{C}) = \{X, Y, Z\}$

Arrows  $\text{hom}_{\mathcal{C}}(X, Y) = X \rightarrow Y$

## Categories, Quickly



Objects  $\text{Ob}(\mathcal{C}) = \{X, Y, Z\}$

Arrows  $\text{hom}_{\mathcal{C}}(X, Y) = X \rightarrow Y$

Composition  $\circ$

# Categories, Quickly

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow^{g \circ f} & \downarrow g \\ & & Z \end{array}$$

Objects  $\text{Ob}(\mathbf{C}) = \{X, Y, Z\}$

Arrows  $\text{hom}_{\mathbf{C}}(X, Y) = X \rightarrow Y$

Composition  $\circ$

Arrows form a monoid under composition

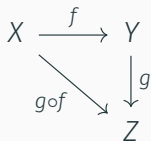
$$\begin{array}{ccccc} W & \xrightarrow{f} & X & & \\ & \searrow^{g \circ f} & \downarrow g & \searrow^{h \circ g} & \\ & & Y & \xrightarrow{h} & Z \end{array}$$

$$(h \circ g) \circ f = h \circ (g \circ f) \quad (3)$$

$$A \curvearrowright id_A$$

$$\forall A. A \in \text{Ob}(\mathbf{C}) \exists id_A : \text{hom}_{\mathbf{C}}(A, A) \quad (4)$$

# Categories, Quickly

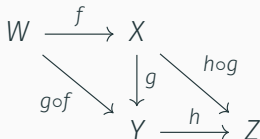


Objects  $\text{Ob}(\mathbf{C}) = \{X, Y, Z\}$

Arrows  $\text{hom}_{\mathbf{C}}(X, Y) = X \rightarrow Y$

Composition  $\circ$

Arrows form a monoid under composition



$$(h \circ g) \circ f = h \circ (g \circ f) \quad (3)$$

$$A \curvearrowright id_A$$

$$\forall A. A \in \text{Ob}(\mathbf{C}) \exists id_A : \text{hom}_{\mathbf{C}}(A, A) \quad (4)$$

## Example

**Set** is the category of sets, where objects are sets, and arrows are functions.

# Functors

The category of (small) categories, **Cat**, has morphisms called Functors.

# Functors

The category of (small) categories, **Cat**, has morphisms called Functors.

These can be thought of as ways to “embed” one category into another.

# Functors

The category of (small) categories, **Cat**, has morphisms called Functors.

These can be thought of as ways to “embed” one category into another.

$$\begin{array}{ccc} \mathbf{FX} & \xrightarrow{Ff} & \mathbf{FY} \\ \uparrow & & \uparrow \\ X & \xrightarrow{f} & Y \end{array}$$

Functors which embed categories into themselves are called Endofunctors.

# Monads

In the category of Endofunctors, **Endo**, a Monad is a triple of:

1. An Endofunctor  $m$ ,
2. A natural transformation:

$$\eta : A \rightarrow m(A) \tag{5}$$

This is an operation which embeds an object.

3. Another natural transformation:

$$\mu : m^2(A) \rightarrow m(A) \tag{6}$$

This collapses two layers of the functor.



# The Category of Measurable Spaces

**Meas** is the category of measurable spaces.

# The Category of Measurable Spaces

**Meas** is the category of measurable spaces.

The arrows ( $\mathbf{hom}_{\mathbf{Meas}}$ ) are measurable maps.

# The Category of Measurable Spaces

**Meas** is the category of measurable spaces.

The arrows ( $\mathbf{hom}_{\mathbf{Meas}}$ ) are measurable maps.

The objects are measurable spaces.

# The Category of Measurable Spaces

**Meas** is the category of measurable spaces.

The arrows ( $\mathbf{hom}_{\mathbf{Meas}}$ ) are measurable maps.

The objects are measurable spaces.

We can construct a functor ( $\mathcal{P}$ ), which, for any given measurable space  $\mathcal{M}$ , is the space of all possible measures on it.

# The Category of Measurable Spaces

**Meas** is the category of measurable spaces.

The arrows ( $\mathbf{hom}_{\mathbf{Meas}}$ ) are measurable maps.

The objects are measurable spaces.

We can construct a functor ( $\mathcal{P}$ ), which, for any given measurable space  $\mathcal{M}$ , is the space of all possible measures on it.

$\mathcal{P}(\mathcal{M})$  is itself a measurable space: measuring is integrating over some variable  $a$  in  $\mathcal{M}$ .

## Implementation<sup>6</sup>

In code (we restrict to measurable functions):

```
newtype Measure a = Measure ((a → ℝ) → ℝ)
```

---

<sup>6</sup>Jared Tobin. *Implementing the Giry Monad*. Feb. 2017. URL: <https://jtobin.io/giry-monad-implementation> (visited on 06/30/2018).

## Implementation<sup>6</sup>

In code (we restrict to measurable functions):

```
newtype Measure a = Measure ((a → ℝ) → ℝ)
```

We also get  $\eta$  and  $\mu$ :

```
integrate :: Measure a → (a → ℝ) → ℝ
```

```
integrate (Measure m) f = m f
```

```
return :: a → Measure a
```

```
return x = Measure (λmeasure → measure x)
```

```
(≫=) :: Measure a → (a → Measure b) → Measure b
```

```
xs ≫= f = Measure (λmeasure → integrate xs
```

```
    (λx → integrate (f x)
```

```
    (λy → measure y)))
```

---

<sup>6</sup>Jared Tobin. *Implementing the Giry Monad*. Feb. 2017. URL: <https://jtobin.io/giry-monad-implementation> (visited on 06/30/2018).

## Other Applications

---



It has been shown<sup>7</sup> that the semantics of the probability monad suitable encapsulate *differential privacy*.

---

<sup>7</sup>Jason Reed and Benjamin C. Pierce. “Distance Makes the Types Grow Stronger: A Calculus for Differential Privacy”. In: *ACM Sigplan Notices*. Vol. 45. ACM, 2010, pp. 157–168. URL: <http://dl.acm.org/citation.cfm?id=1863568> (visited on 03/01/2017).

LINQ<sup>8</sup> is an API which provides a monadic syntax for performing queries (sql, etc.)

PINQ<sup>9</sup> extends this to provide *differentially private* queries.

---

<sup>8</sup>Don Box and Anders Hejlsberg. *LINQ: .NET Language Integrated Query*. en. Feb. 2007. URL:

<https://msdn.microsoft.com/en-us/library/bb308959.aspx>  
(visited on 07/09/2018).

<sup>9</sup>Frank McSherry. “Privacy Integrated Queries”. In: *Communications of the ACM* (Sept. 2010). URL: <https://www.microsoft.com/en-us/research/publication/privacy-integrated-queries-2/>.

# Conclusion

---